

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 11:58:06 ; Search time 32.62 Seconds  
(without alignments)  
3474.464 Million cell updates/sec

Title: US-09-099-898-1  
Perfect score: 453  
Sequence: 1 GTGTGCGGATTGGTTAGC.....GGGAGCTGGAATAAACCT 453

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_GenSeq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	453	100.0	453	1 V84059	CDNA encoding a cy
2	420.6	92.8	553	1 X51720	DNA encoding a hum
3	194	42.8	229	1 X41471	Human secreted pro
4	158.4	35.0	572	1 V84057	CDNA encoding a cy
5	152.4	33.6	560	1 V84056	CDNA encoding a cy
6	92.8	20.5	603	1 V84058	CDNA encoding a cy
7	78	17.2	554	1 V84055	CDNA encoding a cy
8	64.4	14.2	574	1 V84054	CDNA encoding a cy
9	54.4	12.0	527	1 V84053	CDNA encoding a cy
10	40.6	9.0	985	1 V44439	CDNA encoding a cy
11	40.6	9.0	985	1 V64548	Mycobacterium tube
12	39	8.6	8438	1 Q73500	M. tuberculosis im
13	38.6	8.5	114955	1 X53491	Human adenosine Al
14	36.6	8.1	3415	1 T36481	Human integrin bet
15	36.6	8.1	11820	1 V18130	Human chromosome 1
16	36.2	8.0	2932	1 Q25388	TXA2 receptor gene
17	35.2	7.8	329	1 V44425	Mycobacterium tube
18	35.2	7.8	329	1 V44534	M. tuberculosis im
19	34.8	7.7	30001	1 T61016	Total DNA sequence
20	34.8	7.7	117213	1 V62176	HSV-2 strain SB5 C
21	34.8	7.7	30001	1 X05110	S. aureofaciens DN
22	34.8	7.7	114955	1 X53491	Human adenosine Al
23	34.4	7.6	201	1 N70195	Streptomyces prote
24	34.2	7.5	1035	1 T49316	Hypersensitive res
25	34.2	7.5	1035	1 V36430	Hypersensitive res
26	34.2	7.5	1035	1 V39975	Pseudomonas solana
27	34.2	7.5	1035	1 V54609	Pseudomonas solana
28	34.2	7.5	1035	1 V39991	DNA encoding a hyp
29	34	7.5	1687	1 T39798	Human clone 56 gen
30	34	7.5	384	1 V23147	Rat Puralpha like
31	34	7.5	1545	1 V41906	Nucleotide sequenc
32	34	7.5	201	1 X19063	Human PPAR-gamma-1
33	33.8	7.5	6854	1 T76903	S. glaucescens Pst
34	33.8	7.5	535	1 V44428	Mycobacterium tube
35	33.8	7.5	535	1 V64537	M. tuberculosis im
36	33.6	7.4	24379	1 T93095	Streptomyces freno
37	33.6	7.4	24379	1 V25925	Streptomyces roseo
38	33.6	7.4	1439	1 V44424	Mycobacterium tube
39	33.6	7.4	1439	1 V64533	M. tuberculosis im
40	33.6	7.4	1236	1 V72041	Human CP62/R6 vari
41	33.6	7.4	1929	1 V72046	Clone pNG4/A5B7VH-
42	33.6	7.4	1926	1 V72054	806.077 heavy chai
43	33.6	7.4	2019	1 V72059	Plasmid pNG4/55.1s

c 44 33.6 7.4 2025 1 V72064 Fusion protein PNG  
c 45 33.6 7.4 1998 1 V72075 Fusion protein (80

ALIGNMENTS

RESULT 1  
ID V84059 standard; cDNA; 453 BP.  
AC V84059;  
DT 09-MAR-1999 (first entry)  
DE CDNA encoding a cysteine rich soluble protein designated C23.  
KW Cysteine rich soluble protein; CRSP; C23; cell development;  
KW mammalian immune system; antibody; abnormal proliferation; cancer;  
KW inflammation; degeneration; regeneration; atrophy; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 47..373  
FT /tag= a  
FT /product= C23  
FT 47..100  
FT /tag= b  
FT mat\_peptide 101..370  
FT /tag= c  
PN WO9858061-Al.  
PD 23-DEC-1998.  
PR 18-JUN-1998; U12236.  
PR 09-OCT-1997; US-061641.  
PR 19-JUN-1997; US-878730.  
PR 19-JUN-1997; US-878878.  
PR (SCHE ) SCHERING CORP.  
PR Franz-Bacon K, Gorman DM, McClanahan TK;  
PR WPI: 99-095339/08.  
PR P-PSDB; W87710.  
PS New cysteine-rich soluble proteins - used to modulate proliferation,  
PS differentiation, trafficking and development of cells, e.g. for  
PS treating inflammation, cancer and degeneration  
PS Claim 16; Page 18-19; 119pp; English  
CC The present sequence encodes a cysteine rich soluble protein (CRSP)  
CC designated C23. CRSP proteins, and their (ant)agonists, are used to  
CC modulate physiology, differentiation, trafficking and development  
CC of cells (including those in culture), particularly cells of the  
CC mammalian immune system. They are used for treatment of abnormal  
CC proliferation (cancer, inflammation or degeneration), regeneration,  
CC degeneration and atrophy. The proteins are also used to raise, or  
CC detect, antibodies, to design oligonucleotides for library screening,  
CC in drug screens and to isolate cognate receptors. The antibodies are  
CC used for affinity purification of CRSP, to screen expression libraries,  
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
CC competitive drug screens, and as therapeutic modulators.  
SQ Sequence 453 BP; 74 A; 134 C; 158 G; 87 T;

Query Match 100.0%; Score 453; DB 1; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.7e-106;  
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGCGGATTGGTTAGCTAGCCAGGAGCGCTGAGGATGAAGCTCTCTG 60  
|||||  
Db 1 GTGTGCGGATTGGTTAGCTAGCCAGGAGCGCTGAGGATGAAGCTCTCTG 60  
|||||  
QY 61 TCTG 120  
|||||  
Db 61 TCTG 120  
|||||  
QY 121 AGAAGCCATCAATGAGAGGATCCAGGAGTGCCTGCTCCCTAATATTAGGCAATAAG 180  
|||||  
Db 121 AGAAGCCATCAATGAGAGGATCCAGGAGTGCCTGCTCCCTAATATTAGGCAATAAG 180  
|||||  
QY 181 CAGCATTGCTGAGTGCAGGAGCTACCTCAGGGGGAGCTGCTACTTCCCTCCG 240  
|||||

Db 181 CAGCATTTGGCTGGAGTGCAGAGCGCTACCTCCAGGGGGAGCGCTGGCTACTTGGCCCCCG 240  
 QY 241 AGCTTCGCGCGTACCGCGTGCACCTTTGGCTCCGGCTCTGGCTCTGGGATGTGGCGCG 300  
 Db 241 AGCTTCGCGCGTACCGCGTGCACCTTTGGCTCCGGCTCTGGCTCTGGGATGTGGCGCG 300  
 QY 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTGCG 360  
 Db 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTGCG 360  
 QY 361 TGTGACGCGCTGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTGCG 420  
 Db 361 TGTGACGCGCTGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTGCG 420  
 QY 421 CGAGGGGTTGCGGGGAGCTGGAATAAACCT 453  
 Db 421 CGAGGGGTTGCGGGGAGCTGGAATAAACCT 453

RESULT 2  
 ID X51720 standard; DNA; 553 BP.  
 AC X51720:  
 DT 17-JUN-1999 (first entry)  
 DE DNA encoding a human secreted protein.  
 KW Human secreted protein; cancer; immune disorder; infection;  
 KW inflammatory disorder; skin disorder; tumour; atherosclerosis;  
 KW restenosis; autoimmune disorder; Alzheimer's disease;  
 KW peripheral neuropathy; trauma; spinal cord injury; allergy;  
 KW hematopoietic disorder; skeletal disorder; neurological disorder;  
 KW arthritic disorder; asthma; immunodeficiency disease; AIDS;  
 KW transplant rejection; ss.  
 OS Homo sapiens.  
 PN WO91.1293-A1.  
 PD 11-MAR-1999.  
 PF 03-SEP-1997; U18360.  
 PR 12-SEP-1997; US-058974.  
 PR 05-SEP-1997; US-057626.  
 PR 05-SEP-1997; US-057663.  
 PR 05-SEP-1997; US-057669.  
 PR 12-SEP-1997; US-058666.  
 PR 12-SEP-1997; US-058667.  
 PR 12-SEP-1997; US-058973.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Brexer-LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,  
 PI Rosen GA, Ruben SM, Shi Y;  
 DR WPI: 99-204988/17.  
 DR P-PSDB: Y12933.  
 PT New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. neurological disorders,  
 PT tumours, immune disorders, inflammation or haematological disorders,  
 PS Claim 1; Page 165; 215pp; English.  
 CC X51701-55 encode human secreted proteins. The polynucleotides and  
 CC their corresponding secreted polypeptides are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g. by protein or gene  
 CC therapy. Pathological conditions can also be diagnosed by determining  
 CC the amount of the new polypeptides in a sample or by determining the  
 CC presence of mutations in the new polynucleotides. Specific uses are  
 CC described for each polynucleotide, based on which tissues they are  
 CC most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, immune disorders, infection,  
 CC inflammatory disorders, skin disorders, tumours, atherosclerosis,  
 CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral  
 CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic  
 CC disorders, skeletal disorders, neurological disorders, arthritic  
 CC disorders, asthma, immunodeficiency diseases, AIDS and transplant  
 CC rejection. The polypeptides are also useful for identifying their  
 CC binding partners.  
 CC Sequence 553 BP: 160 A; 132 C; 167 G; 94 T;

Query Match 92.8%; Score 420.6; DB 1; Length 553;  
 Best Local Similarity 98.5%; Pred. No. 3e-98;

Matches 446; Conservative 0; Mismatches 4; Indels 3; Gaps 2;  
 QY 1 GTGTGCGGATTTGGTTAGCTGAGCCACGAGAGGCGCTGCAGGATGAAGCTCTCTG 60  
 Db 1 GTGTGCGGATTTGGTTAGCTGAGCCACGAGAGGCGCTGCAGGATGAAGCTCTCTG 60  
 QY 61 TCT 120  
 Db 61 TCT 120  
 QY 121 AGAAGCCATCAATGAGAGGATCAGAGGATCGCGGCTCCCTAATATTTAGGGCAATAAG 180  
 Db 121 AGAAGCCATCAATGAGAGGATCAGAGGATCGCGGCTCCCTAATATTTAGGGCAATAAG 180  
 QY 181 CAGCATTTGGCTGAGTGCAGAGCGTCACTCCAGGGGGAGCGCTGCTACTTCCGCCCG 240  
 Db 181 CAGCATTTGGCTGAGTGCAGAGCGTCACTCCAGGGGGAGCGCTGCTACTTCCGCCCG 240  
 QY 241 AGGCTTCGCGCGTCAACCGGCTGCACCTTGTGGCTCGGCTCTGGGATGTGGCGGC 300  
 Db 241 AGGCTTCGCGCGTCAACCGGCTGCACCTTGTGGCTCGGCTCTGGGATGTGGCGGC 300  
 QY 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTCG 360  
 Db 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGCGCTGCTGTCG 360  
 QY 361 TGTGCGACCGTGGTGGCGCGCGCGCTGCACAGCGCGGGCGGAGCGCGCTCCAGGTC 420  
 Db 361 TGTGCGACCGTGGTGGCGCGCGCGCGCA--GTGGCAACAGCGCGGGCGGCGCGCTCCAGGTC 418  
 QY 421 CGGAGGGGTTGCGGGGAGCTGGAATAAACCT 453  
 Db 419 CGGA-GGGTTGCGGGGAGCTGGAATAAACCT 450

RESULT 3  
 X41471  
 ID X41471 standard; cDNA; 229 BP.  
 AC X41471;  
 DT 22-JUN-1999 (first entry)  
 DE Human secreted protein 5' EST SEQ ID NO: 130 from WO 9906553.  
 KW Human secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductively hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; antinflammatory; tumour inhibition; antitumour; ds.  
 OS Homo sapiens.  
 PN WO9906553-A2.  
 PD 11-FEB-1999.  
 PF 31-JUL-1998; J91237.  
 PR 01-AUG-1997; GS-905051.  
 PA (GEST ) GENSET.  
 DR WPI: 99-153783/13.  
 DR P-PSDB: Y12613.  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from umbilical cord, lymph ganglia,  
 PT lymphocytes and placental tissue  
 PS Claim 1; Page 238; 41pp; English.  
 CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12521 to Y12668,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductively hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, antinflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for

CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 229 BP; 51 A; 58 C; 71 G; 48 T;

Query Match	42.8%	Score 194;	DB 1;	Length 229;
Best Local Similarity	97.5%	Pred. No. 7.4e-41;		
Matches 197;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			

Qy	1	GTGTGCCGGAATTTGGTTAGCTAGACCCACCGAGAGGGCGCTCGAGGATGAAGCTCTCTG	60
Db	28	GTGTGCCGGAATTTGGTTAGCTAGACCCACCGAGAGGGCGCTCGAGGATGAAGCTCTCTG	87
Qy	61	TCTCCTCTCCTCCCTGTCCTGGGGCTGTGTGTCTAGCAGACCTGTGCTCCATGGA	120
Db	88	TCTCCTCTCCTCCCTGTCCTGGGGCTGTGTGTCTAGCAGACCTGTGCTCCATGGA	147
Qy	121	AGAAGCCATCAATGAGAGGATCCAGGAGTTCGCGGCTCCCTAAATATTAGGGCAATTAAG	180
Db	148	AGAAGCCATCAATGAGAGGATCCAGGAGTTCGCGGCTCCCTAAATATTAGGGCAATTAAG	207
Qy	181	CAGATTGGCCTGGAGTGGCAG	202
Db	208	CAGATTGGCCGAGGAGGAG	229

RESULT	4
V84057	
ID	V84057 standard; cDNA; 572 BP.
AC	V84057;
DT	09-MAR-1999 (first entry)
DE	cDNA encoding a cysteine rich soluble protein designated C19.
KW	Cysteine rich soluble protein; CRSP; C19; cell development;
KW	mammalian immune system; antibody; abnormal proliferation; cancer;
KW	inflammation; degeneration; regeneration; degeneration atrophy; ss.

PH	Key	Location/Qualifiers
FT	CDS	29..373
FT		/*tag= a
FT		/product= C19
FT	sig_peptide	29..79
FT		/*tag= b
FT	mat_peptide	80..370
FT		/*tag= c
FT	misc_feature	159..160
FT		/*tag= d
FT		/note= "intron present between these nucleotides in the genomic DNA"
FT	misc_feature	236..237
FT		/*tag= e
FT		/note= "intron present between these nucleotides in the genomic DNA"

PN W0958061-A1.  
PD 23-DEC-1998.  
PF 18-JUN-1998; UL2236.  
PR 19-OCT-1997; US-0611641.  
PR 09-OCT-1997; US-878730.  
PR 19-JUN-1997; US-878730.  
PR 19-JUN-1997; US-878878.  
PA (SCHE ) SCHERING CORP.  
PI Franz-bacon K Gorman DM, McClanahan TK;  
DR WPI: 99-095339/08.  
DR P-PDB: W87708.  
PT New cysteine-rich soluble proteins - used to modulate proliferation,  
PT differentiation, trafficking and development of cells, e.g. for  
PT treating inflammation, cancer and degeneration  
PS Claim 16; Page 16: 11pp; English.  
CC The present sequence encodes a cysteine rich soluble protein (CRSP)  
CC designated C19. CRSP proteins, and their (antagonists, are used to  
CC modulate physiology, differentiation, trafficking and development  
CC of cells (including those in culture), particularly cells of the  
CC mammalian immune system. They are used for treatment of abnormal

CC proliferation (cancer, inflammation or degeneration), regeneration,  
CC degeneration and atrophy. The proteins are also used to raise, or  
CC detect, antibodies, to design oligonucleotides for library screening,  
CC in drug screens and to isolate cognate receptors. The antibodies are  
CC used for affinity purification of CRSP, to screen expression libraries  
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
CC competitive drug screens, and as therapeutic modulators. T.  
SQ Sequence 572 bp: 141 A; 141 C; 166 G; 124 T;

Query Match	35.0%	Score 158.4;	DB 1;
Best Local Similarity	69.2%	Pred. No. 9.4e-32;	
Matches 216;	Conservative	0;	Mismatches 96;
			Indels 0;
			Gaps 0;

Qy	56	CTCTGTCTCCTCCCTCCCTGCTCTGGGCTGTGGTGTCTAGCAAGACCGTGCTGC	115
Db	50	CTCCTTTTCCTTTTCTTCTCTGGGCTGCTGGCCCCAGCATGCTACTGTGTCCC	109
Qy	116	ATGAAGAAGCCATCATGAGAGATCCAGAGATCCCGGCTCCCTAATAATTAGGGCA	175
Db	110	ATGGATGAAGCCATCAGCAAGAGATCAATCAGACTTCAGCTCCCTACTGCCAGTGCA	169
Qy	176	ATAAGCAGCATGGCTTGGAGTCCACAGAGCGTCACTCCAGGGGGGACCCTGGCTACTTGC	235
Db	170	ATGAAGAACACTGTCTACATTGCTGTGCTACTCTCCAGAGGGAGGCTGGGCTCCTCG	229
Qy	236	CCCCGAGCCTCGCGCTCACCGGCTGACATGTGGCTCCGCCCTGTGCTCGTGGGATGTG	295
Db	230	CCAGAAGGCACAACCGTCATAGCTGCTCTCTGTGGCTCTGGCTGTGGCTCATGGACGTC	289
Qy	296	CGCGCCGAGCCACATGCTCACTGCCAGTGGCGGGCATGGACTGGACCGGAGCGCGTGC	355
Db	290	CGTGAGGATACAATGTCTCACTGCCAGTGGCGGAAGCATAGACTGGACAGGCGGCCGTGC	349
Qy	356	TGTCGTGTGCAG	367
Db	350	TGTACCTGTGGG	361

```

RESULT      5
V84056
ID          V84056 standard; cDNA; 560 BP.
AC          V84056;
DT          09-MAR-1999 (first entry)
DE          cDNA encoding a cysteine rich soluble protein designated Cl9.
KW          Cysteine rich soluble protein; CRSP; Cl9; cell development;
KW          mammalian immune system; antibody; abnormal proliferation; cancer;
KW          inflammation; degeneration; regeneration; degeneration; atrophy; ss
OS          Mus sp.
FH          Key
FT          Location/Qualifiers
FT          CDS               64..408
FT          FT               /*tag= a
FT          FT               /product= Cl9
FT          sig_peptide      64..123
FT          FT               /*tag= b
FT          mat_peptide      124..405
FT          FT               /*tag= c
FT          misc_feature      193..194
FT          FT               /*tag= d
FT          FT               /note= "intron present between these nucleotides in
FT          FT               the genomic DNA"
FT          misc_feature      271..272
FT          FT               /*tag= e
FT          FT               /note= "intron present between these nucleotides in
FT          FT               the genomic DNA"
PN          W09858061-A1.
PD          23-DEC-1998.
PF          18-JUN-1998; U12236.
PR          09-OCT-1997; US-061641.
PR          19-JUN-1997; US-878730.
PR          19-JUN-1997; US-878878.
PA          (SCHE ) SCHERING CORP.

```





```
RESULT 9
V84053
ID V84053 standard; cDNA; 527 BP.
AC V84053;
DT 09-MAR-1999 (first entry)
DE cDNA encoding a cysteine rich soluble protein designated C2.
KW Cysteine rich soluble protein; CRSP; C2; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 32..367
FT FT /*tag= a
FT FT /product= C2
FT FT sig_peptide 32..100
FT FT /*tag= b
FT FT mat_peptide 101..364
FT FT /*tag= c
PN W09858061-Al.
PD 23-DEC-1998.
PF 18-JUN-1998; U12236.
PR 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE ) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McClanahan TK;
DR WPI: 99-095339/08.
DR P-PSDB; W87704.
PT New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiation, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 16; Page 12-13; 119pp; English.
CC The present sequence encodes a cysteine rich soluble protein (CRSP)
CC designated C2. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise, or
CC detect, antibodies, to design oligonucleotides for library screening,
CC in drug screens and to isolate cognate receptors. The antibodies are
CC used for affinity purification of CRSP, to screen expression libraries,
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC to produce anti-idiotypic antibodies (useful for diagnosis), in
CC competitive drug screens, and as therapeutic modulators.
SQ Sequence 527 BP; 135 A; 126 C; 116 G; 150 T;

Query Match 12.0%; Score 54.4; DB 1; Length 527;
Best Local Similarity 57.7%; Pred. No. 2.2e-05;
Matches 97; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 197 TGCAGAGCGTCACTCCAGGGGGACCTGGCTACTTGGCCCCCGAGGCTTCGCCGTACC 256
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 TGCAGTAGTGTCAAGACTATGAACAGATGGGCTCTCTGCCCTCTGGGATGACTGCTACT 253
QY 257 GCGTGCACCTTGCGTCCGCTGCTGGTGGATGTGGCCGCCAGACACATGTCAAC 316
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GGGTGTGCTGTGGCTTGGCTGTGGATCTTGGGAGATCCAGAGTGGAGATCTTGCAC 313
QY 317 TGCAGTGGCGGCGATGAGTGGAGCGGCGCTGCTGTGCTGTG 364
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TGCCTGTGCTTACTCGTTGACTGGACACTGCCCCGCTGCTGCCAATG 361

RESULT 10
V44439
ID V44439 standard; DNA; 985 BP.
AC V44439;
DT 09-NOV-1998 (first entry)
DE Mycobacterium tuberculosis antigen XP25 3'-end DNA.
KW Mycobacterium tuberculosis; infection; diagnosis; XP25; ss.

Mycobacterium tuberculosis strain Erdman.
OS W09816645-A2.
PN 23-APR-1998.
PF 07-OCT-1997; U18214.
PR 13-MAR-1997; US-818111.
PR 11-OCT-1996; US-729622.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR WPI: 98-251292/22.
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis.
PS Claim 4; Page 187-188; 250pp; English.
CC This is the 3' region of DNA coding for an antigenic portion of
CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
CC V44438. XP25 DNA was isolated from a M. tuberculosis strain Erdman
CC genomic DNA expression library using sera from patients having
CC extrapulmonary tuberculosis. It bears no similarity to known
CC sequences. The invention relates to methods and compositions for
CC diagnosing tuberculosis. It provides polypeptides (see
CC W64291-W64379) comprising an antigenic portion of a soluble M.
CC tuberculosis antigen, or an immunogenic portion of a M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.
SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match 9.0%; Score 40.6; DB 1; Length 985;
Best Local Similarity 45.3%; Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 111 GCTCCATGAAGAAGCATCATGAGAGATCCAGGAGTCCGGCGTCCCTATATTTA 170
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 GCGCCGCGGTGCGAGGGCAAGGGCGGCGGAGGTGCGGGGTCCGGCGCGACA 157
QY 171 GGCATATAGCAGCATTTGGCTTGAGTGCAGAGCGTCACTCCAGGGGGACCTGGCTA 230
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 ACCCCACCGTGTGTTTCCGCGGTGGCGCGCACAGGTGGCGCGCGCGCGCG 217
QY 231 CTTGCCCGCAGGCTTCGCGGTGCACGCGTGCCTTGTGCTCGCTGCTGCTGG 290
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 GCGGGCGCGCGGGGCGGACCGGTACCGCGCGCACCGCGGCGTGTGCGGCCACCGGTA 277
QY 291 ATGTGCGCGCGAGACACACATGTCACTGCGCGGGGCGATGGACTGGACCGAGCGC 350
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 GTGAGGATCGCGGGGCGCGCGCGCGCGGTGACGCGGGCGATGGGCGCGGTC 337
QY 351 GCTGCTGTGTGTGACCCCTGAGGTGCGCGCACAGCGGTGCACAGCGCGGGGAGCGC 410
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 TCGGCTGTGGCCCTCTCCGCGTGTGACGCGCGCAAGGCGGCGGCGCGCGCGCA 397
QY 411 GCTCCAGTCCGAGGGGTTGCCGGGG 437
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 GCGCGCGCGCGCGCGCATCAACGGGG 424

RESULT 11
V64548
ID V64548 standard; DNA; 985 BP.
AC V64548;
DT 27-JAN-1999 (first entry)
DE M. tuberculosis immunogenic polypeptide XP25 3'-end DNA.
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis; ss.
OS Mycobacterium tuberculosis.
PN W09816646-A2.
DR 23-APR-1998.
PF 07-OCT-1997; U18293.
```





PF 17-SEP-1998; U19419.  
PR 09-JUN-1998; US-093972.  
PR 17-SEP-1997; US-059160.  
PA (UYEC-) UNIV EAST CAROLINA.  
PI NYCE JW;  
DR WPI; 98-229400/19.  
PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
PT vasoconstriction  
PS Disclosure; Page 37; 120pp; English.  
CC The specification describes antisense oligonucleotides (X52869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene  
CC initiation codons, genomic flanking regions, intron-exon borders, the  
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
CC regions and all segments of RNAs encoding proteins associated with one  
CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
CC may be derived from sequences X55272-74. These multiple target  
CC oligonucleotides (specifically X55180-271) can be used for the antisense  
CC treatment of diseases and conditions. Typical diseases and conditions  
CC are those associated with impaired respiration and inflammation,  
CC including lung diseases, pulmonary vasoconstriction, inflammation,  
CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,  
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary  
CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic  
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic  
CC metastases, as well as all types of cancers which may metastasize or have  
CC metastasized to the lungs, including breast and prostate cancer.  
SQ Metastasis 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;  
Query Match 8.5%; Score 38.6; DB 1; Length 114955;  
Best Local Similarity 36.0%; Pred. No. 0.65;  
Matches 93; Conservative 19; Mismatches 146; Indels 0; Gaps 0;  
QY 180 GCAGCATTCGCTGAGTCCAGAGCGTCCACCTCCAGGGGGGACCTGGCTACTTGGCCCCC 239  
Db 103818 GCGCGCTCGGGCGGCSNNNDNCGTCBTGGCGCGTGGCGCGGCSNNNDNCGTC 103877  
QY 240 GAGGCTTCGCCCTACCGCTGACCTGTTGGTTCGCCCTGTTGGCTGTTGGATGTCGCGC 299  
Db 103878 BTGGCGGCTCGGGCGGCSNNNDNNGTCBTGGCGCGTGGCGCGGCSNNNDNNTCB 103937  
QY 300 CCGAGACCATCTCACTCCAGTGGCGGGGATGACCTGGACCGGACGCGCTGCTGTC 359  
Db 103938 TGGCGGCTCGGGCGGCSNNNDNNCBTGGCGGCTGGCGGCGGCSNNNDNBTGGCG 103997  
QY 360 GTGTGAGCCCTGAGTTCGCGGCGAGCGGTGCACAGCGCGGCGGCGGCTCCAGGT 419  
Db 103998 GCGTCGGCGGCSNNNDNNTGTCGGCGCTGGCGCGGCSNNNDNNGCGGCTCGGC 104057  
QY 420 CCGAGAGGTTGCGGGG 437  
Db 104058 CCGGCSNNNDNNGCGCG 104075  
RESULT 14  
T36481  
ID T36481 standard; cDNA; 3415 BP.  
AC T36481;  
DE 08-OCT-1996 (first entry)  
DT Human integrin beta subunit protein, beta-5, cDNA.  
KW Human; integrin beta subunit; beta-5; carcinoma; lymphoid cell;  
KW immunoassay; detection; mRNA; assay; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 337..2736  
FT /tag= a  
FT signal\_peptide 337..405  
FT /tag= b  
FT mat\_peptide 406..2733  
FT /tag= c

PN US5527679-A.  
PD 18-JUN-1996.  
PR 01-MAY-1991; 694314.  
PR 01-MAY-1991; US-694314.  
PR 27-APR-1993; US-054077.  
PA (DAND ) DANA FARBER CANCER INST INC.  
PI Hemler ME, Ramaswamy H;  
DR WPI; 96-299852/30.  
DR P-PSDB; W02194.  
PT Immunoassay and mRNA hybridisation assay for beta-5 protein - useful  
PT for the detection of carcinoma(s) and to distinguish different cell  
PT types  
PS Claim 11; Columns 13-20; 21pp; English.  
CC The present sequence encodes the human integrin beta subunit  
CC protein, beta-5, which is found in carcinomas but not in lymphoid  
CC cells. An immunoassay for the detection of beta-5, comprises  
CC contacting a sample with a monoclonal antibody (Ab) which binds 1  
CC epitope of beta-5, and then with a labelled Ab which binds another  
CC epitope of beta-5 and detecting any bound label. An assay for  
CC beta-5 mRNA, comprises contacting a sample with a probe capable of  
CC hybridising to the beta-5 cDNA, and determining if binding has  
CC occurred. These assays are useful for detecting carcinomas, and for  
CC distinguishing between different cell types.  
SQ Sequence 3415 BP; 752 A; 943 C; 1009 G; 711 T;  
Query Match 8.1%; Score 36.6; DB 1; Length 3415;  
Best Local Similarity 47.6%; Pred. No. 1;  
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;  
QY 140 ATCCAGGAGTCCCGCGCTCCCTTAATATTAGGGCAATAAGCAGCATTCGCTGAGTGC 199  
Db 1690 AGCTCGAGGTGGGGTGCACCTACACTGCACGTGGCGTGGCGGTGGGACCC 1749  
QY 200 CAGAGCGTCACCTCCAGGGGGGACCTGGCTACTTGGCCCGAGGCTTCGCCGTACCGGC 259  
Db 1750 AACAGCGCCAGGTGCAACGGGAGCGGACCTATGTCTGCGGCTGTGTAGTGAGCGCC 1809  
QY 260 TGCACCTTGGTCCGCTGTGGCTGTGGGATGTGGCGCCGAGACCATGTCATGTC 319  
Db 1810 GGCTACCTGGGACCGAGTCCGAGTCCAGGATGGGAGAACGAGCGTGTACCAAGAC 1869  
QY 320 CAGTGGCGGCGATGATGACGACCGGAGCGCTGTGTGCTGTGCA 366  
Db 1870 CTGTGGCGGAGGACAGGACACCTGTGACAGCGGCGTGGGA 1916  
RESULT 15  
V18130/C  
ID V18130 standard; DNA; 11820 BP.  
AC V18130;  
DT 04-SEP-1998 (first entry)  
DE Human chromosome 19 derived USF2 gene sequence.  
KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;  
KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;  
KW multicystic renal dysplasia; renal agenesis; hydronephrosis;  
KW Von Mayer-Rokitansky-Kuester disorder; bifid ureter; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT 5'UTR 1..1088  
FT /tag= a  
FT CDS 1089..11189  
FT /tag= b  
FT /product= "USF2 gene product"  
FT /note= "contains introns"  
FT exon 1089..11150  
FT /tag= c  
FT /number= 1  
FT intron 1151..1435  
FT /tag= d  
FT /number= 1  
FT exon 1436..1482  
FT /tag= e



FT	/number= 2
FT	.1483..1571
FT	/*tag= f
FT	/number= 2
FT	.1572..1690
FT	/*tag= g
FT	/number= 3
FT	.1691..1792
FT	/*tag= h
FT	/number= 3
FT	.1793..1993
FT	/*tag= i
FT	/number= 4
FT	.1994..2437
FT	/*tag= j
FT	/number= 4
FT	.2438..3588
FT	/*tag= k
FT	/number= 5
FT	.3589..2708
FT	/*tag= l
FT	/number= 5
FT	.2709..2796
FT	/*tag= m
FT	/number= 6
FT	.2797..3074
FT	/*tag= n
FT	/number= 6
FT	.3075..3132
FT	/*tag= o
FT	/number= 7
FT	.3133..10631
FT	/*tag= p
FT	/number= 7
FT	.10632..10726
FT	/*tag= q
FT	/number= 8
FT	.10727..10879
FT	/*tag= r
FT	/number= 8
FT	.10880..11008
FT	/*tag= s
FT	/number= 9
FT	.11009..11099
FT	/*tag= t
FT	/number= 9
FT	.11100..11186
FT	/*tag= u
FT	/number= 10
FT	.11190..11820
FT	/*tag= v
FT	.11723..11728
FT	/*tag= w

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 11:58:03 ; Search time 164.82 Seconds  
(without alignments)  
5421.415 Million cell updates/sec

Title: US-09-099-898-1  
Perfect score: 453  
Sequence: 1 GTGTGGCGGATTGGTTAGC.....GGGGAGCTGGAAATAACCT 453

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST :

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*
- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*

54: em\_est22:\*

55: em\_est23:\*

56: em\_est24:\*

57: em\_est25:\*

58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	422.4	93.2	477	31	AA3111223	AA311223 EST181980
2	175.2	38.7	194	24	N41594	N41594 yw66h04.r1
3	152.4	33.6	480	38	AA796548	AA796548 vbl4f11.r
4	152.4	33.6	530	47	A1527863	A1527863 u13f10.y
5	152.4	33.6	535	47	A1528963	A1528963 u43e05.y
6	150.8	33.3	484	38	AA796118	AA796118 v06e08.r
7	150.8	33.3	508	39	AA823170	AA823170 v04e10.r
8	143.4	31.7	432	42	A1121371	A1121371 uc30h11.r
9	140.4	31.0	493	29	AA139082	AA139082 mr11g09.r
10	136.2	30.1	379	26	W42057	W42057 mb16d01.r1
11	134	29.6	496	41	A1021019	A1021019 ua99a04.r
12	132.8	29.3	372	26	W42069	W42069 mb16d04.r1
13	132	29.1	445	33	AA423511	AA423511 v876s03.r
14	131	28.9	371	43	A1180917	A1180917 ub77c04.r
15	129.6	28.6	511	34	AA467519	AA467519 ve01d11.r
16	128	28.3	385	44	A1286748	A1286748 ub96a04.r
17	127.4	28.1	402	28	AA097250	AA097250 mk16a08.r
18	124.2	27.4	472	38	AA795022	AA795022 v808e05.r
19	123	27.2	453	41	A1036421	A1036421 ub64a02.r
20	120.2	26.5	520	34	AA472183	AA472183 v998d06.r
21	104.2	23.0	227	41	A1021285	A1021285 ub07a10.r
22	98.8	21.8	271	40	AA980745	AA980745 ua45b05.s
23	96.6	21.3	309	42	A1120050	A1120050 uc26d09.r
24	96.2	21.2	422	27	W34958	W34958 mc34f06.r1
25	91.8	20.3	419	31	AA286376	AA286376 vc48c04.r
26	89.6	19.8	290	37	AA718169	AA718169 vu54h05.r
27	88.2	19.5	577	34	AA524300	AA524300 q32q12.s
28	85.2	18.8	403	49	AV011470	AV011470 AV011470
29	79	17.4	513	37	AA692356	AA692356 vt58f01.r
30	79	17.4	407	37	AA734659	AA734659 vr90b11.r
31	78	17.2	487	37	AA711012	AA711012 vt52h07.r
32	77.8	17.2	586	36	AA615920	AA615920 vo91d01.r
33	77.8	17.1	388	37	AA691748	AA691748 vs13h05.r
34	77.4	17.1	484	37	AA711094	AA711094 vt55e11.r
35	77.2	17.0	389	30	AA273994	AA273994 vb70e06.r
36	76.6	16.9	319	49	AV021952	AV021952 AV021952
37	76.4	16.9	527	30	AA245583	AA245583 my52d07.r
38	76.2	16.8	319	39	AA823296	AA823296 vp37a09.r
39	75.8	16.7	626	34	AA518288	AA518288 vi27h06.r
40	75.6	16.7	469	37	AA711284	AA711284 vt71c12.r
41	75.4	16.6	490	37	AA689884	AA689884 vt62f12.r
42	75.2	16.6	408	35	AA592737	AA592737 vo37a09.r
43	74.2	16.4	522	36	AA638656	AA638656 vo55h09.r
44	74	16.3	524	39	AA822133	AA822133 vp35c03.r
45	73.4	16.2	354	36	AA638525	AA638525 vo54c10.r

ALIGNMENTS

RESULT 1

AA311223

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

AA311223 477 bp mRNA EST 19-APR-1997

EST181980 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.

AA311223

g1963551

AA311223.1 GI:1963551

EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 477)  
 REFERENCE  
 AUTHORS Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A., Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D., White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C., Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D., Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geoghagen N.S., Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr., Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M., Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M., Phillips C.A., Ryder S.E., Scott J.L., Saudke D.M., Shirley R., Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y., Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J., Dinke D., Feng D.-F., Ferrie A., Fischer C., Hastings G.A., He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K., Kozak D.L., Kunsch C., Hungjun J., Li H., Meissner P.S., Olsen H., Raymond L., Wei X.F., Wang J., Xu C., Yu G.L., Ruben S.M., Dillion P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C., Fraser C.M. and Venter J.C.  
 TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 96026280  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1398072.  
 Other\_ESTs: THC144666  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerl@vetigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tldb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..477  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):157191"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells v"  
 /cell\_type="T-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 81 a 132 c 167 g 94 t 3 others  
 ORIGIN

Db 241 AGGCTTCGGCGTNCACGGCTGCACCTGTGGCTCCGCGCTGTGGATGTGGCG 300  
 QY 300 CCAGACACCATCTCACTGCAGTGGCGGCGCATGGACTGGAGCGGCGGCTGCTGTC 359  
 Db 301 CCAGACACCATCTCACTGCAGTGGCGGCGCATGGACTGGAGCGGCGGCTGCTGTC 360  
 QY 360 GTGTGAGCCCTGAGTTCGGCGGCGCATGGACTGGAGCGGCGGCTGCTGTCAGGT 419  
 Db 361 GTGTGAGCCCTGAGTTCGGCGGCGCATGGACTGGAGCGGCGGCTGCTGTCAGGT 417  
 QY 420 CCGAGGGGTTGGCGGCGGAGCTGGAATAAACCT 453  
 Db 418 CCGAGGGGTTGGCGGCGGAGCTGGAATAAACCT 451

RESULT 2  
 N41594 194 bp mRNA EST 24-JAN-1996  
 DEFINITION YW66H04.r1 Soares\_placenta\_8to9weeks\_2NbHP8to9w Homo sapiens cDNA  
 clone IMAGE:257239 5', mRNA sequence.  
 ACCESSION N41594  
 NID 91185625  
 VERSION N41594.1 GI:1165625  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 194)  
 AUTHORS Hillier L., Clark N., Dubucq T., Elliston K., Hawkins M., Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F., Trevisakis E., Waterston R., Williamson A., Wohlmann P. and Wilson R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On May 9, 1995 this sequence version replaced gi:802396.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 High quality sequence stops: 174  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: T7  
 High quality sequence stop: 174.  
 Location/Qualifiers  
 1..194  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3886849"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Soares\_placenta\_8to9weeks\_2NbHP8to9w"  
 /dev\_stage="two placenta: one from 8 weeks and another from 9 weeks post conception"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCGGCGGATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."  
 BASE COUNT 30 a 60 c 39 t 3 others  
 ORIGIN

Query Match 93.2%; Score 422.4; DB 31; Length 477;  
 Best Local Similarity 98.5%; Pred. No. 1.5e-88;  
 Matches 447; Conservative 0; Mismatches 3; Indels 4; Gaps 2;  
 QY 1 GTGTGCGGATTTGGTTAGCTGAGCCACCAGAGGCGGCTGCAGGATGAAGCTCTCTG 60  
 Db 1 GTGTGCGGATTTGGTTAGCTGAGCCACCAGAGGCGGCTGCAGGATGAAGCTCTCTG 60  
 QY 61 TCTG 120  
 Db 61 TCTG 120  
 QY 121 AGAAGCCATCAATGAGAGGATCAGAGGATCGCGGCTCCCTATATTTAGGGCAATAAG 180  
 Db 121 AGAAGCCATCAATGAGAGGATCAGAGGATCGCGGCTCCCTATATTTAGGGCAATAAG 180  
 QY 181 CAGCATTTGGCTGGAGTGCAGAGCGTACCTCCAGGGGGACCTGGCTACTTCCGCCCG 240  
 Db 181 CAGCATTTGGCTGGAGTGCAGAGCGTACCTCCAGGGGGACCTGGCTACTTCCGCCCG 240  
 241 AGGCTTCGGCGT-CACCGGCTGCATTTGGCTCCGCGCTGTGGCTGTGGATGTGGCG 299

T	3'				
<p>daptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo.*</p>					
BASE COUNT	107 a	133 c	121 g	119 t	
ORIGIN					
Query Match	33.6%	Score 152.4;	DB 38;	Length 480;	
Best Local Similarity	67.8%;	Pred.No.	3.6e-26;		
Matches 213; Conservative	0;	Mismatches 101;	Indels 0;	Gaps 0;	
QY	54	CTCTGTGTCCTCCCTCCTCCCCTGTCTCGGGGGTGTTGGTGTCTAGCAAGACCCCTGTGCT	113		
Dd	87	CCTCTCTTTCTTTCTTTCTTTCTTGCCTGAATGCTGTGGCTGCCAGCATGCCACTGTGTC	146		
QY	114	CCATGGAAGAAGCCATCAATGACGAGATCCAGGAGTGC GGCGGTCCCTAATAATTATFAGG	173		
Dd	147	CCATCGATGAAGCCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAATG	206		
QY	174	CAATAAGCACATTTGGCTCGGAGTGCCAGAGGTCACCTCCAGGGGGACCTGGCTACTT	233		
Dd	207	CAATAAGAACATTTGGCTTAATTTCTGGACAGTCTCTCCAGAGGAAGTTGGCCTCT	266		
QY	234	GCCTCCGAGGCTTCGCGCTCACGGCTGCACCTTTGGCTCGGCTGTGGCTCGTGGGATG	293		
Dd	267	GCCAGRAGGCACAGCAGTCTTAGCTGCTCCTGTGGCTGTGCTGTGGCTGTGGGACA	326		
QY	294	TGGCGCGCGAGACACATGTCATGACCAGTCGCGCGGCGATGGATGACCGAGGCGGCT	353		
Dd	327	TTCGTGAAGAAAAAGTGTGCTACTGCCAGTGTGCAAGGATAGACTGGACAGCAGCCGCT	386		
QY	354	GCTGTGCTGTGCAG	367		
Dd	387	GCTGAAGCTGCAG	400		
RESULT	4				
AISZ7663					
LOCUS	AISZ7663	530 bp	mRNA	EST	18-MAR-1999
DEFINITION	uf13fio.yl Soares mouse mammary gland NLMKG Mus musculus cdna clone IMAGE:1511275 5', mRNA sequence.				
ACCESSION	AISZ7663				
NID	94441798				
VERSION	AISZ7663.1	GI:4441798			
KEYWORDS	EST.				
SOURCE	tissue mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 530)				
AUTHORS	Marra,M., Hillier,B., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons.M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.				
TITLE	The WashU-NCI Mouse EST Project 1999				
JOURNAL	Unpublished (1999)				
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3187318.  Contact: Marria M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mousetest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="#">info@image.llnl.gov</a> ) for further information. MGI:938127 Seq primer: -40RP from GIBCO High quality sequence stop: 481.				



[illegible]





Fax: 314 286 1810  
Email: mouse@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1362616  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 448.

## FEATURES

```

1. /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone_image="IMAGE:597184"
   /clone_lib="Soares mouse 3NbMS"
   /sex="male"
   /tissue_type="Spleen"
   /dev_stage="4 weeks"
   /lab_host="DH10B"
   /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; left strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTTACCAATCGAAGTGGGAGCGGCCTGTATTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
110 a    139 c    122 g    122 t

```

BASE COUNT  
ORIGIN

Query Match	31.0%	Score 140.4;	DB 29;	Length 493;
Best Local Similarity	67.5%;	Pred. No. 2.2e-23;		
Matches 212; Conservative	0;	Mismatches 101;	Indels 1;	Gaps 1;

[illegible]

RESULT 10

LOCUS	W42057	379 bp	mRNA	EST	11-SEP-1996
DEFINITION	mb16d01.rl Soares mouse p3NMF19.5 Mus musculus cDNA clone				
IMAGE:	329569 5', mRNA sequence.				

ACCESSION	W42057
NID	g1326540
VERSION	W42057.1
KEYWORDS	GI:1326540
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa;

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Importance of Teacher Professionalism	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Effect of Teacher Salary on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 379)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kuback, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Apr 14, 1993 this sequence version replaced gi:716841.

Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:210969

Seq primer: ETPrimer	
High quality sequence stop: 373.	
Location/Qualifiers	
source	1. .379

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="IMAGE:329589"
/clone_lib="Soares mouse p3NMFL9.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker: Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAAGTCGGACGGCGCGATATTTTATTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaído. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

91 a 104 g 95 t

BASE COUNT

Query Match 30.1%; Score 136.2; DB 26; Length 379;  
Best Local Similarity 66.6%; Pred. No. 1.9e-22;

QY	54	CTCTCTGTC	TCCTCCTCCTCCCTGTCCTGGGGCTGTTGGTGTCTAGCAAGACCCCTGGCT	113
Db	87	CCCTCCTCTTCTCTTCTCTCTGTCCTCGAAGCTGCTGGCTCCAGCATGCCATCGTC	146	
QY	114	CCATGGAAGAAGCCATCAATGAGAGATCCAGGAGGTCGCGGCTCCCTAATATTTAGGG	173	
Db	147	CCATCGATGAAGCCATCGACAAGAAGTCAACGACTTCACTCCTCTTTTCCAAATG	206	
QY	174	CAATTAAGCAGCATGGCTCGAGTGCCAGAGCGTCACTCCAGGGGGAGCATGGCTACTT	233	
Db	207	CAATAAAGAACATTTGGCTTAAATTGCTGGACAGTCTCCTCCAGAGGGAAGTTGGCCCTCT	266	
QY	234	GCCCCGAGGCTTCGCGGTACACGGCTGCACTGTGGCTCCGCTGTGGCTCGCTGGGATG	293	
Db	267	GCCCAAGAGGCACAGCAGTCTTAGCTGTCTCTGTGGCTCTGCTGTGGCTGGGACA	326	
QY	294	TGCGCGCCGAGACCATGTCACTGCCAGTGCAGCGGCGCATGGACTGGACCGGA	346	
Db	327	TTCCGTGAAGAAAAGTGTGTCACTGCCAGTGTCAAGGATAGACTGTGACAGCA	379	

RESULT 11  
AI021019

LOCUS	AI021019	496 bp	mrna	EST	16-JUN-1998
DEFINITION	ua99a04.rl Soares mouse mammary gland NbMMG Mus musculus cDNA clone IMAGE:1365582 5', mRNA sequence.				
ACCESSION	AI021019				
NID	9323535				
VERSION	AI021019.1	GI:3235355			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 496) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE	The WashU-HMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2151315.				
FEATURES	<p>Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:898802 Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 458. Location/Qualifiers 1..496 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /map="21q" /clone="IMAGE:1365582" /clone_lib="Soares mouse mammary gland NbMMG" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B" /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I-oligo(dT) primer [5', TGTTCAACTCTCAAGTGGGCGGCGGAATGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."</p>				
BASE COUNT	109 a	140 c	124 g	123 t	
ORIGIN	<p>Query Match 29.6%; Score 134; DB 41; Length 496; Best Local Similarity 67.0%; Pred. No. 6.5e-22; Matches 205; Conservative 0; Mismatches 100; Indels 1; Gaps 1;</p>				
QY	54	CTCTCTGTCCTCCTCCTCCTCCTGTCCTGCGGCTGTGGTGTCTAGCAAGACCCGTGTGCT	113		
Db	87	CCCTCCATTTCTCTTTCTCTGTCCTGAACCTGCTGGGTCCA-CATGCCACTGTGTC	145		
QY	114	CCATGGAGAGGCATCAATGAGAGGATCCAGAGGTCGCCGGTCCCTATATTAGGG	173		
Db	146	CCATCGATGAAGCCATCGACAAAGATGAACAAAGACTTCAACTCCCTGTTTCCAAATG	205		
QY	174	CAATAAGCAGCATTTGGCCTGGAGTGCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTT	233		

/note="Organ: mammary gland; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGGGCGCCGAAATGGTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      102 a    120 c    113 g   ~ 110 t  
ORIGIN

Query Match                  29.1%;    Score 132;    DB 33;    Length 445;  
Best Local Similarity       65.9%;    Pred. No. 1.8e-21;  
Matches 207; Conservative    0; Mismatches 105; Indels    2; Gaps    1;

QY 54 CTCCTGTCTCCTCCTCCCTCCTCCTCCTGCTGTGGGGTGTTGGTGTCTAGCAAGACCCTGTGCT 113  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 91 CCCCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 150  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 114 CCATGGAGAAGCATCAATGAGAGATCCAGAGGTCGCCGGCTCCCTTAATTATTTAGG 173  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 151 CCATCGATGAAGCCATCGCAAGAAGATCAACAAGACTTCAACTCCCTGTTTCCAATG 210  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 174 CAATAAGCAGATTGGCTTGAGTGCCAGCGTFCACCTCCAGGGGGAGCTGGCTACTT 233  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 211 CAAATAAGACATTGGCTTAATTTCTGGACAGTCTCTCCAGAGGAAGTTGCCTCT 270  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 234 GCCCCGAGGCTTCGCCGTCCACGGCTGCATTTGTGTGCTCCGCTGTGGCTGTGGGATG 293  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 271 GCCAAGAAGCACAGCAGTCTTGAGCTGCTCCTGTGCTCTGC--TGTGCTGTGGGACA 328  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 294 TGGCGCGGAGACACATGTCATCTGCAGTCGCCGGCATGGACTGACCGGAGCGCGCT 353  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 329 TTCGTGAAGAAAAGTGTGCTCACTGCCAGTGTGCAAGATAGACTGGACAGACGCCGCT 388  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 354 GCTGCTGTGTGCAG 367  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 389 GCTGTAAGCTGCAG 402  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14

AI180917

LOCUS AI180917 371 bp mRNA EST 08-OCT-1998

DEFINITION ub77c04.rl Soares mouse mammary gland NMLMG Mus musculus cDNA clone IMAGE:1383750 5', mRNA sequence.

ACCESSION AI180917

NID G3731555

VERSION AI180917.1 GI:3731555

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1407470.

Contact: Marria M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 11:58:04 ; Search time 144.5 Seconds  
(without alignments)  
9970.105 Million cell updates/sec

Title: US-09-099-898-1  
Perfect score: 453  
Sequence: 1 GTGTGCGGATTTGGTTAGC.....GGGAGCTGGAATAAACCT 453

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pil.\*
- 8: gb\_p12.\*
- 9: gb\_prl.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_st.\*
- 14: gb\_sts.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_vl.\*
- 18: em\_fun.\*
- 19: em\_htg.\*
- 20: em\_hum1.\*
- 21: em\_hum2.\*
- 22: em\_in.\*
- 23: em\_on.\*
- 24: em\_or.\*
- 25: em\_ov.\*
- 26: em\_pat.\*
- 27: em\_ph.\*
- 28: em\_pl.\*
- 29: em\_ro.\*
- 30: em\_sy.\*
- 31: em\_un.\*
- 32: em\_un.\*
- 33: em\_vl.\*
- 34: gb\_htg1.\*
- 35: gb\_htg2.\*
- 36: gb\_in1.\*
- 37: gb\_in2.\*
- 38: em\_bal.\*
- 39: em\_ba2.\*
- 40: em\_hum3.\*
- 41: em\_hum4.\*
- 42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
1	43.8	9.7	981 12	RNCKERIMP X54806 R.norvegicu

2	43.4	9.6	41055	1	SC6G4	AL031317 Streptomy
3	41.6	9.2	40883	11	AC005776	AC005776 Homo sapi
4	40.8	9.0	39441	11	AC005565	AC005565 Homo sapi
5	40.8	9.0	5180	12	RNU48596	U48596 Rattus norv
6	40.6	9.0	47852	1	MTV023	AL022022 Mycobacte
7	40.2	8.9	47852	1	MTV023	AL022022 Mycobacte
8	40	8.8	5880	17	HSE4	D1486 Equine herp
9	39.8	8.8	7082	1	SFU08223	U08223 Streptomyce
10	39.8	8.8	1320	3	CVCMYC	X95367 C.familiari
11	39.4	8.7	806	7	HVU49482	U49482 Hordeum vul
12	39.4	8.7	2578	12	MUSHOXA	L08757 Mus musculu
13	39.2	8.7	2157	4	XELBLAA	M6363 X.borealis
14	39.2	8.7	3252	10	HSTAF113	Y11354 H.sapiens m
15	39.2	8.7	42189	11	AC005258	AC005258 Homo sapi
16	39	8.6	129354	11	AC004080	AC004080 Homo sapi
17	39	8.6	2691	11	AF040714	AF040714 Homo sapi
18	39	8.6	3637	12	AF045017	AF045017 Mus muscu
19	39	8.6	3579	12	MMAJ3298	AJ223298 Mus muscu
20	39	8.6	5123	17	HEPVIE	X15120 Pseudorabie
21	39	8.6	8438	17	SHILLT	M57505 Pseudorabie
22	38.8	8.6	907	7	HVGRP3	248624 H.vulgare m
23	38.8	8.6	170030	11	AC005829	AC005829 Homo sapi
24	38.6	8.5	4246	11	AF055376	AF055376 Homo sapi
25	38.6	8.5	2145	11	AF055377	AF055377 Homo sapi
26	38.6	8.5	2579	12	MUSMDR2A	M74151 Mouse p-gly
27	38.4	8.5	3375	1	AMPEPSYNT	X97860 Amycolatops
28	38.4	8.5	40790	1	MTCY493	295844 Mycobacteri
29	38.2	8.4	2427	9	BABINTB5GL	L12231 Yellow babo
30	38.2	8.4	24254	10	HSAC001228	AC001228 244Kb Con
31	38.2	8.4	4000	12	MUSPOUDOMA	M88299 Mouse brain
32	38	8.4	67200	1	MTV017	AL021897 Mycobacte
33	37.8	8.3	36583	1	SC5H1	AL049863 Streptomy
34	37.8	8.3	785	8	AF010580	AF010580 Oryza sat
35	37.8	8.3	5721	12	RNCA171	278279 R.norvegicu
36	37.6	8.3	164296	34	HSAC000380	AC000380 Homo sapi
37	37.4	8.3	84364	9	HS222J7	298885 Human DNA s
38	37.4	8.3	35465	11	AC005525	AC005525 Homo sapi
39	37.4	8.3	110000	34	HSDJ689N3_1	Continuation (2 of
40	37.4	8.3	3372	42	AF065164	AF065164 Homo sapi
41	37.2	8.2	7515	11	HSU26644	U26644 Human fatty
42	37	8.2	43832	1	SC6A5	AL049485 Streptomy
43	36.8	8.1	2028	8	SS1132828	AJ132828 Spermatoz
44	36.8	8.1	5225	10	HSU06154	U06154 Human clone
45	36.8	8.1	1756	10	HUMD2A	M30625 Human dopam

ALIGNMENTS

RESULT	1	RNCKERIMP	981 bp	RNA	ROD	30-APR-1992
LOCUS		R.norvegicus	cyto			
DEFINITION		R.norvegicus	cyto			
ACCESSION		X54806				
NID		955955				
VERSION		X54806.1				
KEYWORDS		cyto				
SOURCE		Norway rat.				
ORGANISM		Rattus norvegicus				
REFERENCE		1 (bases 1 to 981)				
AUTHORS		Redfern,C.P.F.				
TITLE		Direct Submission				
JOURNAL		Submitted (26-SEP-1990) Redfern C.P.F., Univ. of Newcastle upon Tyne, Medical Mol. Biology Group, Dept. of Dermatology, 4th Floor Cookson Bldg. Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK				
REFERENCE		2 (bases 1 to 981)				
AUTHORS		Redfern,C.P.F. and Allen,G.				
TITLE		Sequence variation (between species) at the carboxyterminal domain of cyto				
JOURNAL		Unpublished				







```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 40883)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,
Burkhardt-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Dol., Regala,W., Terry,A., Gaines,J.,
Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trong,S.,
Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 1.5 Mb OLFR-rich region in 19p13.1
Unpublished
2 (bases 1 to 40883)
Lamerdin,J.E.
Direct Submission
Submitted (05-OCT-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from p telomere to centromere. Cosmid
R28550 overlaps cosmid R28194 (AC003111) to the left from bases 1
to 504 of this accession, and is separated from cosmid R33907 to
the right by a sequence gap of <2 kb. Additional chromosome 19 map
and sequence information are available at:
http://www-bio.lnlnl.gov/dbp/genome/genome.html.
FEATURES
Location/Qualifiers
1..40883
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R28550"
/chromosome="19"
/map="19p13.1 between OLFR and D19S885"
/cell_line="5HL2-B"
/clone_lib="LL19NC02 F chromosome 19 specific cosmid
library"
/note="Cosmid library constructed at LNLL from flow-sorted
chromosomes from human-hamster hybrid 5HL2-B, which
carries chromosome 19 as its only human chromosome."
repeat_region 1..245
/rpt_family="AluSq"
repeat_region 246..442
/rpt_family="LIMB8"
repeat_region 443..754
/rpt_family="AluSq"
repeat_region 770..963
/rpt_family="LIMB1"
repeat_region 1161..1282
/rpt_family="AluSq/x"
repeat_region 1286..1585
/rpt_family="AluSq"
repeat_region 1588..1880
/rpt_family="AluSq"
repeat_region 1891..2012
/rpt_family="AluJb"
repeat_region 2024..2326
/rpt_family="AluY"
repeat_region 2327..2616
/rpt_family="AluSq"
repeat_region 2776..2841
/rpt_family="LINE2"
repeat_region 3495..3622
/rpt_family="LINE2"
repeat_region complement(3667..3766)
/rpt_family="MER3"
misc_feature complement(4287..4396)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
repeat_region 4507..4787
/rpt_family="AluSq"
repeat_region 4808..5109
/rpt_family="AluY"
repeat_region 5283..5545
/rpt_family="AluSq"
6064..6207
/rpt_family="MIR"
repeat_region 6537..6577
/rpt_family="(TGG)n"
repeat_region 8122..8401
/rpt_family="AluJb"
repeat_region 8402..8565
/rpt_family="AluJb/FRAM"
repeat_region 8585..8889
/rpt_family="AluJb"
repeat_region 9886..10177
/rpt_family="AluSx"
repeat_region 10179..10251
/rpt_family="(GAAA)n"
repeat_region complement(11205..11323)
/rpt_family="AluY"
repeat_region 11363..11494
/rpt_family="AluY"
repeat_region 11506..11820
/rpt_family="AluY"
repeat_region 11821..12003
/rpt_family="AluYa5"
repeat_region 12022..12128
/rpt_family="FLAM_A"
repeat_region complement(12200..12259)
/rpt_family="MIR"
repeat_region complement(12910..13002)
/rpt_family="MIR"
misc_feature 13871..14036
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 83.000"
repeat_region complement(14095..14209)
/rpt_family="MIR"
repeat_region 14562..14951
/rpt_family="AluSp"
repeat_region complement(15085..15108)
/rpt_family="AT-rich"
repeat_region 15109..15431
/rpt_family="AluSx"
misc_feature 15662..15770
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 80.000"
repeat_region 16107..16190
/rpt_family="LIMB8"
repeat_region 16191..16497
/rpt_family="AluSx"
repeat_region 16498..16937
/rpt_family="LIMB8"
repeat_region complement(17602..17708)
/rpt_family="LINE2"
repeat_region complement(17709..18005)
/rpt_family="AluSp"
repeat_region 18393..18690
/rpt_family="AluSc"
misc_feature 18706..18968
/note="BLASTN similarity to 741380 (105..368); match:
0.99, score: 3.8e-100; database searched: est;
phib12_19/1TV Homo sapiens cDNA clone phib12_19/1TV."
repeat_region 19418..19703
/rpt_family="AluSx"
repeat_region complement(20667..20801)
/rpt_family="GC-rich"
repeat_region complement(20802..20829)
/rpt_family="(GGA)n"
repeat_region complement(20871..20955)
/rpt_family="(CGGG)n"
repeat_region complement(20982..21134)
/rpt_family="GC-rich"
repeat_region complement(21237..21352)
/rpt_family="(CGGG)n"
repeat_region complement(21349..21470)
/rpt_family="(CGG)n"
repeat_region 21482..21559

```





QY	355	CTGTCTGTGTGAGCCCTTGCAGCGGCGAGCGGTCTGCACAGCCCGGCGGAGCGCGCTC	414
Db	15058	CAGCCCGCGGAGCCCTTGGAGACACAGCGCCGCTGTCTCTGTGAGCGGCTGCTC	15117
RESULT	5		
RNA48596/c			
LOCUS	5180 bp	RNA	05-JUN-1996
DEFINITION	Rattus norvegicus MAP kinase kinase 1 (MEKK1) mRNA, complete cds.	ROD	
ACCESSION	U48596		
NID	gl354136		
VERSION	U48596.1		
KEYWORDS	GI:1354136		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 5180)		
AUTHORS	Xu, S., Robbins, D.J., Christerson, L.B., English, J.M., Vanderbilt, C.A. and Cobb, M.H.		
TITLE	Cloning of rat MEK kinase 1 cDNA reveals an endogenous membrane-associated 195-kDa protein with a large regulatory domain		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (11), 5291-5295 (1996)		
MEDLINE	9622476		
REFERENCE	2 (bases 1 to 5180)		
AUTHORS	Cobb, M.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-FEB-1996) Melanie H. Cobb, Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75235, USA		
FEATURES	Location/Qualifiers		
source	1..5180		
	/organism="Rattus norvegicus"		
	/db_xref="taxon:10116"		
gene	516..4997		
	/gene="MEK1"		
CDS	516..4997		
	/gene="MEK1"		
	/codon_start=1		
	/product="MAP kinase kinase 1"		
	/protein_id="AAC52596.1"		
	/db_xref="PID:gl354137"		
	/db_xref="GI:1354137"		
	/translation="MAAAGDRASSGFFGPAASAPGAGGGGALQGGAPAGAGL RETSGAGRADWRQOLRRSHVDELQPEQLFLTASPCCSTSPSPPEPADAAAGA SGFQPAAGCPGCAAGCSHSAELAAARDGARGSPAGAPPSAASPAGSGREMNKFTL KGLHKDDRPPEERIREKLCATCPANKHEWLRNRNRCPPVVKPIPIKDGSEMENL AALQEGQAGNAPKAPGRSPSPSGRSGKSGVRNRRVSPVPFGSGNLP PRAPSPGFSPSPETSRVKNMRLYLQQIPNSFLIGGDSPNKYRVFTIP QNCSGRGFTCHLLFVMLRVFOLESPDMLRWKTLKNEFVSLFKYHSRRSRKKA PSRNTIQKVRSMNGHTLSSSTSSSENSTKDEEQMPTCLGLMLDESLTVKE DGCNRKLHHCKSIWAEECPRRNEPLICPLCRSKWRSHDFYSHELSPVDSPSTSLRGV QQPSPQQPVAGSQRNRQNSFNLTHTGQIPPAIKDLAEPIQAFGMELVSGCLFSR NMNREMARLHNSVSGALLNATHSTGTGGGGSLSAGAASSQSPSISGDVFE AFCSVLISACADPYVYVYAAALTKRLMVLPTCHSLERIKLQRLIRGLVGLNKA ADANSTLSISLTLELCKGAGELAVGREILKAGSIGVGVYVYVLSICILVGNQASFN NQWELGLCLIDLRLLEISAEFYPHIVSTDVSAQEPVEIRYKLLSLAFALQSIDN SHSWGKLSRRYILSRRAVTVPPPLFSKLVMSLSSGSHFAMRRRLMAIADEVEI AEVIQGSEDTLDGQSDSAAVTPPFPRESSLEHTAHVEKTKGKLKATLSASSDI SDRLAGVGLSPSSATPEPKPTVQKGRPHSQCLNSPLSPQLMFPFPAISAPCSSAP SVPAQSVDAKSHRPAFVQSTIPSAQSPQKRFLOQFCTCSDNSDKLSVPFTQBS RPPPSNIRHAKASRPVPGSTSKLGDAKNSMTLDNSASOCDSDFGSGNSGSAVIP SETATPTADKCRLDVNPENLSIEDLLEASMPSDTTVTYKSEVAVLSPKEASDI TYKDDVNNQCKKEKMEEEEEALAMAMASQDALPVPVQLVENGEDILIIQDIT PEFLPHTKANREYREDTEWLKQQLGAFSGSQAOVDGVTGLMAVKQVTVYRNTS SOEEVEYAEALREIRMSHNLNHNIIIRMLGATCEKSNYLFIEWMAASVAHLLSYK APKESVYINTEOLLRGLSVLHENOIIHRDVKGANLLIDSTGORLFIADFGAARLAS KGTGAGEFGQGLLGTAFNAPVLRGQQYGRSCDVMWSGVQADRIEMAKCPPWNAEKHS NHUALLFKASATAPSPHSLGSDVLRDLVALRCLELQDQRRPSELLKHPVFTTW"		
BASE COUNT	1225 a	1521 c	1443 g 991 t
ORIGIN			

Query Match	9.0%	Score 40.8	DB 12	Length 5180
Best Local Similarity	50.0%	Pred. No. 4.5		
Matches 102	Conservative 0	Mismatches 102	Indels 0	Gaps 0
QY	201	AGAGCGTCACCTCCAGGGGACCTGGCTACTTGCCTCCGAGGCTTCGCCGTACCGGCT	260	
Db	607	AGAGTCGCCCGCCGCCCGCTCGGGGCTCGCCGCGCGCGGCCGCGGGAATCCGAC	548	
QY	261	GCATCTGTGGCTCGCCCTGTGCTCGTGGGATGTGGCGCCGAGACCATCTCACTGCC	320	
Db	547	GACGAGGCGCATCGCCGCGCGCCGCCCATTTCTTCGGGCGTACATTCGTCGTG	488	
QY	321	AGTGGCGGGCATGACTGACGAGCGCGCTGCTGCTGTGTCAGCCCTGAGTTCGGG	380	
Db	487	CCTCGGGGAGGAGGGGGGAGAGCGCGGGGGCGGCGGAGCGGGCGCGG	428	
QY	381	CGCAGCGCGTCCACGCGCGGGCG	404	
Db	427	CGCTGCGCCGAGTGCGGGGCGG	404	
RESULT	6			
MTV023				
LOCUS	MTV023	47852 bp	DNA	BCT
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.			17-JUN-1998
ACCESSION	AL022022	AL123456		
NID	93261554			
VERSION	AL022022.1	GI:3261554		
KEYWORDS	Mycobacterium tuberculosis.			
SOURCE	Mycobacterium tuberculosis			
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
REFERENCE	1 (bases 1 to 47852)			
AUTHORS	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.			
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence			
JOURNAL	Nature 393 (6685), 537-544 (1998)			
MEDLINE	98295987			
REMARK	Erratum:[published erratum appears in Nature 1998 Nov 12;396(6707):190]]			
REFERENCE	2 (bases 1 to 47852)			
AUTHORS	Parkhill, J.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk			
COMMENT	On Jun 27, 1998 this sequence version replaced gi:2924430.			

Notes: Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.  
**CAUTION:** In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome

binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

## source

Location/Qualifiers  
1. 47852

/organism="Mycobacterium tuberculosis"

/strain="H37Rv"

/db\_xref="taxon:1773"

<1. 18

/organism="Mycobacterium tuberculosis"

/strain="H37Rv"

/db\_xref="taxon:1773"

/clone="Y13E12"

complement(1..1695)

/gene="Rv3494c"

/note="Rv3494c. (MTV023.01c), len: 564. Unknown protein similar to several Mycobacterium tuberculosis proteins e.g. MTC128.14 (515 aa), MTC19H5.28c (516 aa) and (MTV051.09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: 297050|MTC128.14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 297050|MTC19H5.28 (516 aa) opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in 555 aa overlap. Tbpase score is 0.897"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein Rv3494c"

/protein\_id="CAAL7731.1"

/db\_xref="PID:el254622"

/db\_xref="PID:g2924431"

/db\_xref="GI:2924431"

/db\_xref="SPTREMBL:O53539"

/translation="MIDRLAKIQLSIFAVITVITLSVMAIFVLRLPATFEGTGYGVS  
DFVAGGGLKKNANVYRGVAVGVSVGLNPGVTAHRLNSGTAIPSNVATVRSVS  
AIGQYIDLVPENSSKLRNGFRQKQNRIGQDVADLLRQAETLLSGDGLKRLR  
LSHAEIATNGAPPEALRIEASRLVDEANANPVQSOLIDQAGFFLQAQIRAGGDI  
KSLADGAFRTWQLRAADPLRLADAPDAIDEANTAFSGIRPFPALAAASLANLGR  
VGVYHKSLEQLLVVEPFAALITTSAGGVODEGAKLDFKIDLDPPTCMGTGELPPP  
LVRSADPESRIPDMVCKTAQNDPSTVRGARNYPCQEEFGKRAFTVQLCRDPRGYV  
PVGTNWRGPPPIPYTEVDGRNLLPNKFFIIPGADPDGCVIVGPPPGQVAGPG  
PAPHPAPQAPGPNNGPPPTSMWPPGYPPPPVYPATIPPPPPPEGGPPPGP  
APGPOQASGPATYTDLSGAFADPAGGTGIFAPGTCASSAENWVDMRDPRLQ"

complement(1..1695)

/gene="Rv3494c"

<1. 47852

/note="fragment designated v023. Does not represent a

physical clone"

complement(1703..1707)

/note="possible RBS for Rv3494c"

complement(1706..2860)

/gene="lprN"

complement(1706..2860)

/gene="Rv3495c"

/note="Rv3495c. (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTC19H5.29 (402 aa) and (MTV051.08). Probably lipoprotein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: gp|297050|MTC128.13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and

297182|MTC19H5.29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. Tbpase score is 0.897"

/codon\_start=1

/transl\_table=11

/product="lprN"

/protein\_id="CAAL7732.1"

/db\_xref="PID:el254623"

/db\_xref="PID:g2924432"

/db\_xref="GI:2924432"

/db\_xref="SPTREMBL:O53540"

/translation="MNRILWRAILITASSALLAGCGGGLNSILPLPGTAGHGAYSV

## misc\_feature

/gene="lprN"

/note="PS00013 Prokaryotic membrane lipoprotein lipid

attachment site"

complement(2857..4212)

/gene="Rv3496c"

complement(2857..4212)

/gene="Rv3496c"

/note="Rv3496c. (MTV023.03c), len: 451. Unknown but

similar to Mycobacterium tuberculosis proteins MTC128.12 (530 aa), MTC19H5.30c (508 aa) (MTV051.07). Hydrophobic region at N-terminus. FASTA scores: 297050|MTC128.12 (530 aa) opt: 838 z-score: 711.0 E(): 4.3e-32; 35.1% identity in 473 aa overlap; and 297182|MTC19H5.30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. Tbpase score is 0.891"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein Rv3496c"

/protein\_id="CAAL7733.1"

/db\_xref="PID:el254624"

/db\_xref="PID:g2924433"

/db\_xref="GI:2924433"

/db\_xref="SPTREMBL:O53541"

/translation="NMGRVAMLTGKRYATVIALVAALVGGVYVLSSTNKRITVG  
YFTSAVLGPGDQVGLVGVGEIDMIEPRSDVKITMSVDKVPVDVQAVIMSPN  
LYAARFIQLTPYITGGAVLPDNGRIDLDTATVPVDEVEKEGLTRLAADLSPAGELQ  
GPELAAINQAADTLGNGDSLHNLARELAQVAGRLGDSRGDIFGTVKNLQVLDALSE  
SDEQIVQAGHVSQVLSADLANLDLGTINQALSDIRGLFRRNNSTLIETVNLQ  
NDFATLSQSENIQVHLVAGPTNFYNIYDPAQTGLISLINFANPVQFIGCG  
SFTDAGSPADPYRRAECIRERLGVRLTVNYPIMFHLPLNTITAYKG011YDTP  
ATEAKSETVPPELTWVPAGGAPVGNPADLSLLVPPAPAPAPAGPAGPEHGCG  
G"

complement(2867..2871)

/gene="Rv3496c"

/note="possible RBS for Rv3495c"

complement(4209..5282)

/gene="Rv3497c"

complement(4209..5282)

/gene="Rv3497c"

/note="Rv3497c. (MTV023.04c), len: 357. Unknown but

similar to Mycobacterium tuberculosis proteins MTC19H5.31 (481 aa), MTC128.11, (515 aa) and MTV051.06. Hydrophobic region at N-terminus. FASTA scores: 297182|MTC19H5.31 (481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 32.5% identity in 332aa overlap; and 297050|MTC128.11 (515 aa) opt: 587 z-score: 641.2 E(): 3.3e-28; 30.1% identity in 335 aa overlap. Tbpase score is 0.889"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein Rv3497c"

/protein\_id="CAAL7734.1"

/db\_xref="PID:el254625"

/db\_xref="PID:g2924434"

/db\_xref="GI:2924434"

/db\_xref="SPTREMBL:O53542"

/translation="MLNRKPSKHERDPLRTIGFLVLCVVLIAFGYSGLPFWPG  
KYDAYFTDAGGITPGNSVYGLKVGSAVSAVLAGNSAKVTFSDRSIVVQGS  
LAAITDTTLGHSITAVSPAGSKSTIPLSTRTTPTTLNGLVQLDGRANDLNRRQFEQA  
LNVT0ALHDATPOVRGAVDGLTSLSRALNRDEALQGLLAHAKSVTSVLSRAEQVN  
KLVEDGNOLFALDARPAALSAISIGDIDVAAQISGFVADNRKRFEPALSKNLVLAN  
LMNRDYLITEALKRLTYATTIGVEVVGSGPGFNVNIVSLVPLGLVATVFLVFPQGLK  
PDSLADYLRGFTQERWIIRKPS"

complement(5272..6324)

/gene="Rv3498c"

complement(5272..6324)

/gene="Rv3498c"



```
0; 33.6% identity in 363 aa overlap; and
297182|MTCI19H5_29 (402 aa) opt: 572 z-score: 667.9 E():
1.le-29; 31.8% identity in 362 aa overlap. Tbpaise
score is 0.897
/codon_start=1
/transl_table=11
/product="lprn"
/protein_id="CAAL7732.1"
/db_xref="PID:el254623"
/db_xref="PID:g2924432"
/db_xref="GI:2924432"
/db_xref="SPTREMBL:O53540"
/translation="MNRILWRAIIITASSALLAGQCFGLNSLPLPGTAGHGEYASV
KVEDAVATLQNSVPMVDVTVGSVAGIVAQVRDGSFYAARLKLDKNLLPANAV
AKVSTSLGSLHVELAPDRPTPTGLVDGSRITEANTDREPTTEEVFSALGVVYVK
GNVGALEIIDEHOAVAGROAFVNLPRLAELTAGLNROVHDIIDALDGLNRYSAI
LARQDNLGRALDTLPDAVRVNLQNRHDHVDFAALKRLTWTSHVLAETKVDGFEDL
KDLISVKNALNDRDKDFTVSLQLLTFEFPNFGIKOAVRGDYLNVFTFDLRLRIGE
TFFTAYDFDNNAHDEILNPPDFLIGELIANLSGOAADPFKIPPGTASQ"
/complement(2798..2830)
/gene="lprn"
/Note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
/complement(2857..4212)
/gene="Rv3496c"
/complement(2857..4212)
/gene="Rv3496c"
/Note="Rv3496c. (MTV023.03c), len: 451. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.12
(530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
region at N-terminus. FASTA scores: 297050|MTCI28.12 (530
aa) opt:838 z-score: 711.0 E(): 4.3e-32; 35.1% identity
in 473 aa overlap; and 297182|MTCI19H5.30 (508 aa) opt:
821 z-score:697.0 E(): 2.6e-31; 35.1% identity in 453 aa
overlap.Tbpaise score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3496c"
/protein_id="CAAL7733.1"
/db_xref="PID:el254624"
/db_xref="PID:g2924433"
/db_xref="GI:2924433"
/db_xref="SPTREMBL:O53541"
/translation="MMGRVAMLTGSGRLRYATVIALVAGVGVVLSSTGNKRTIVG
YFTSAGVLPQDOVVLGVPGVEIDMIEPRSSDVKITSVSKDVFPVDVQAVMSPN
LVAAFIQLTPTVTCGAVLPDNGRIDLDRKTAVPVEWDEVEKGLTRLAADLSPAAGELQ
GLPAAIINQAADTLQNGDSLHNALRELAQVAGRLGDSRGDIFGTAKNQLVLDALSE
SDEQIVFAGHVASVSVQLADSANLQDTLGLNQALSDIRGFLRENNSLTLETYNQL
NDFQTLSDSENIEQVLHVAGPITNFYNIYDPAQGTGLNGLSLPFPANPVOFICGG
SFDTAAGSPADYVRAEICRERLGPVLRRLTVNYPPIFMFPLNTITAYKGOIIVDTP
ATEAKSEFPVPLTVPAGGAPVGNPDLQSLVPPAPGAPAPAGPAGGGEHGG
G"
/complement(2867..2871)
/gene="Rv3496c"
/Note="possible RBS for Rv3495c"
/complement(4209..5282)
/gene="Rv3497c"
/complement(4209..5282)
/gene="Rv3497c"
/Note="Rv3497c. (MTV023.04c), len: 357. Unknown but
similar to Mycobacterium tuberculosis proteins MTCY19H5.31
(481 aa), MTC128.11, (515 aa)and MTV051.06. Hydrophobic
region atN-terminus. FASTA scores: 297182|MTCI19H5_31
(481 aa) opt:611 z-score: 667.7 E(): 1.le-29; 32.5%
identity in 332aa overlap; and 297050|MTCI28.11 (515 aa)
opt: 587 z-score:641.2 E(): 3.3e-28; 30.1% identity in
335 aa overlap. Tbpaise score is 0.889"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3497c"
/protein_id="CAAL7734.1"
/db_xref="PID:el254625"
/db_xref="PID:g2924434"
```

```
/db_xref="GI:2924434"
/db_xref="SPTREMBL:O53542"
/translation="MLNRKPSKHERDPLRTGIFGLVLVICVVLIAFGYSGLPFWQOG
KTYDAVFTDAGGITPGNSVYVSLKVGAVSAVSLAGNSAKVTFSDVRSIVVGQSLAA
IRTDIILGERSIATVPAGSGKSTTTPSLRTTPTTLNGVLQDLGRNANLNRQFEQA
LNVFTQALHATPQVQVAGVDTLSLRALNRDEALQGLLAHAKSVLSLSRAEQVN
KLVEDGNOLFALDARRAALSALISGIDDDVAAQISGFVADNRKEFGPALSKNLVLAN
LNERRDYITEALKRLTPTATTTLGVEVYVSGPGFNVNYSVLPGLVATVDFLVPQGL
PDSLADYLRGFIQERWIIRPKSP"
/complement(5272..6324)
/gene="Rv3498c"
/complement(5272..6324)
/gene="Rv3498c"
/Note="Rv3498c. (MTV023.05c), len: 350. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.10
(346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
region at N-terminus. FASTA scores: 297050|MTCI28.10 (346
aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in
Query Match
Best Local Similarity 8.9%; Score 40.2; DB 1; Length 47852;
Matches 78; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 203 ACGGTACCTCCAGGGGACCTGGCTACTTGCCCGGAGCTTCGCCGCGCTGC 262
Db 31717 ACGGTGCGCCTTGCGCGCGCGCCACCTTTCCGGCGCGCGCGCGCGCGC 31658
QY 263 ACTTGTGGCTCCGCCCTGTGGCTGCTGGGATGTCGGCGCGAGACCATGTC 322
Db 31657 CCGCGCGGTGCGCGCGCTGATATTGCGCGCGCTTCGCCGCGCGCGCGC 31598
QY 323 TCGCGGGGATGACTGGACC 343
Db 31597 ACCGCGGACACCTTGACC 31577

RESULT 8
HSE4 HSE4 5880 bp DNA VRL 03-FEB-1999
LOCUS Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and
DEFINITION glycoprotein H (gH) genes.
ACCESSION D14486 D00683 D00684
NID 9221820
VERSION D14486.1 GI:221820
KEYWORDS UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
SOURCE Equine herpesvirus 4 (strain 1942).
ORGANISM Equine herpesvirus 4
VIRUSES; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE 1 (sites)
AUTHORS Nicolson,L., Cullinane,A.A. and Onions,D.E.
TITLE The nucleotide sequence of an equine herpesvirus 4 gene homologue
of the herpes simplex virus 1 glycoprotein H gene
J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
90362066
2 (sites)
REFERENCE 2 (sites)
AUTHORS Nicolson,L., Cullinane,A.A. and Onions,D.E.
TITLE The nucleotide sequence of the equine herpesvirus 4 thymidine
kinase gene
J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
90362067
3 (bases 1 to 5880)
REFERENCE 3 (bases 1 to 5880)
AUTHORS Nicolson,L.
JOURNAL Unpublished (1993)
MEDLINE Submitted (10-SEP-1990) to DBJ by:
REFERENCE 4
AUTHORS Lesley Nicolson
JOURNAL Dept. Veterinary Pathology. University of Glasgow
COMMENT Vet School
Bearsden Roda, Glasgow G61 1QH. Scotland
UK.
Phone: 041-339-8855
Fax: 041-330-5733.
Location/Qualifiers
FEATURES
```



```
source          1. .5880
                /organism="Equine herpesvirus 4"
                /db_xref="taxon:10331"
polyA_signal    /note="put. polyadenylation signal (UL24); putative"
                complement(99. .104)
polyA_signal    /note="put. polyadenylation signal (UL24); putative"
                complement(115. .120)
CDS             /note="put. polyadenylation signal (UL24); putative"
                complement(116. .934)
                /note="putative; ORF1 (UL24 homologue)"
                /codon_start=1
                /protein_id="BAA03377.1"
                /db_xref="PID:d1003887"
                /db_xref="PID:g221821"
                /db_xref="GI:221821"
                /translation="MKRKQRLTARSLRAGIRCHNRFNAMVODLASAKKNVYCARL
                APFSELPVATLKTAMGVSLAFVNLGQRRPDCVCTVQFHGSDAGKVCILILKTC
                RFSKNMTASKNLKGMROLHDSCLLAKTLPGSGEILLAPLVFAVQGRMRLR
                VTRLSPOVYSNAVLSTISRLAEYSPPISESTRRCTVTRTNSAFRAKTTGSI
                OPTQAKPATANAASLFSATAQANTNAVGYOPATISLANLAWASLPAK"
                861. .865
TATA_signal     /note="put. TATA box (TK); putative"
                887
misc_feature     /note="pot. transcription initiation site; putative"
                949. .2007
CDS             /partial
                /note="ORF2; (TK)"
                /codon_start=1
                /protein_id="BAA03378.1"
                /db_xref="PID:d1003888"
                /db_xref="PID:g221822"
                /db_xref="GI:221822"
                /translation="MAACVPPEGEAPRSAGTPTTRQVTIVYILDGVYIGKSTGTVR
                SAAASGGSTLTPPEPMAYWTFETDVISGIDYDNRKQGVSTVCFDAALITAHQY
                MRFTTPLYLLDHPCTLFGNSLQRTQPDLTLYDHRPVASTVCFPAARVLLGDMSM
                CALMAVATLTPREGQGNIVYTLNVEHRRRLTRARIGRIDITLIATIRNYFHL
                VNTCHFRSGRWKDGDELPTSCGAYKHRATQMDAFQERVSPELGDTLFLFKTQEL
                LDDRGVLYEHWALADKLKRLNLVFSADLSGTPTQCAAVVESLLPLMSTLSDFD
                SASALERAARTFNAEMGV"
                2049. .2054
polyA_signal    /note="put. polyadenylation signal (TK); putative"
                2127. .2133
TATA_signal     /note="putative"
                2197. .2202
TATA_signal     /note="putative"
                2225. .4792
CDS             /partial
                /note="ORF3; (gH)"
                /codon_start=1
                /protein_id="BAA03379.1"
                /db_xref="PID:d1003889"
                /db_xref="PID:g221823"
                /db_xref="GI:221823"
                /translation="MSOPYLKIALVIAIVSIPVMTPTVSTPQOQKLHYVNGT
                WNNHTNFTYRIDTMEPVYNNLSSTFEVAISERNFTVNTPLGASVFWLKSL
                NPKHQPCIANPPEGPRGPNVSTVSLFNDLEPLMTKNLEFVLPDNTITGW
                TFEBSKVYAKNGVGVVSPPTSPDNNITRDDGTPKQHLSTIIDHTTFVLDLQNE
                TKTLTISPRNAVPIATFAGITVMGCDTQAIAYLNGPWGQIQISVNNPPELMIV
                APNDVRARIVNLRPRRLEPPGPTAGTIYKIVYVLSQGNFYLGHSKISREVAAYPE
                ESDYRHLSLANDTLAMLAELSNGSKSVSYLYRIARLAVATFSLAEVIRLSQY
                MLEQADIMVLELIVPLNVKAAAGTADSSYSSVDAMDQFVAQAIEKIVADIN
                IENELRPMYEHRSLLVNVAYSRKPLPNVAFNRLITAMYKEAIKDRITWNTMRE
                VLFVAGAAAGSHVILDDGPDGLHAKDSMFLSLNRNILLCTAMCTASHAYSAGV
                KLEVMAGLISAGVQVSLLEVFSPCMASAREFDLAEERVLDSLLVPRRLTDLNTGL
                EDDGTTTHSVGRSINLSRINAYNFDAVRFTPELASCSTKLPKVLVPLANRSY
                VITRPNIGLTYSLDGVNIAKPIVISYITYGNCQVSRAITRSVLDHPGHTQSCYVC
                GSVFRMYASGAIMDLTYIDDKQVLEQLVAGENSTIPAFNPKLYTPSNMALLMPNGT
                VTLMSPASYSAPKIPSTYLMWASIGLLAILIYIVVKMCGGVINNDYSLLNSE"
                4777. .4782
polyA_signal    /note="put. polyadenylation signal(gH); putative"
                4893. .5315
repeat_region   /note="direct repeat region"
                5395. .5403
rep_origin      /note="put. replication origin; putative"
```

```
rep_origin      5440. .5448
                /note="put. replication origin; putative"
                complement(5755. .5880)
                /note="putative; ORF4 (UL21 homologue)"
                /codon_start=1
                /protein_id="BAA03380.1"
                /db_xref="PID:d1003890"
                /db_xref="PID:g221824"
                /db_xref="GI:221824"
                /translation="NVSILITAFDNLDRKGAUYLYLLSERLKRGGVPVHVNR"
BASE COUNT     1459 a 1377 c 1499 g 1545 t
ORIGIN
Query Match    8.8%; Score 40; DB 17; Length 5880;
Best Local Similarity 47.3%; Pred. No. 6.6;
Matches 121; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 180 GCAGCATTCGCTGAGTCCAGAGCGTCACTCCAGGGGGGACCTGGCTACTTGCCTCC 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5004 GCGGCGCGCGCGGAGGCTGCGGCGCGCGCGGAGGCTGCTGCGGCGCGGCGGAGG 5063
QY 240 GAGGCTTCGCTCACCGGCTGCACTTGTGGCTCCGCTGTGGCTGTGGATGTGCGCG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5064 GCTGTGCGGCGCGCGGAGGCTGCTGCGGCGCGCGGAGGCTGCTGCGGCGCGGCGG 5123
QY 300 CCGAGACCATGTCTACTCCAGTCGCGGCGGATGACCTGGACCGGAGCGCGTGTGTC 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5124 AGGCTGTGCGGCGCGCGGAGGCTGCTGCGGCGCGCGGAGGCTGCTGCGGCGCGG 5183
QY 360 GTGTGACGCTTCAAGTCCGCGCGACGCGGTGCACAGCGCGCGGCGGCGGCTCCAGGT 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5184 GCGAGGCTGCTGCGGCGCGCGGAGGCTGCTGCTGCGGCGCGGCGGAGGCTGCTGCT 5243
QY 420 CCGGAGGGTTCGGGG 435
    ||| ||| ||| |||
Db 5244 GCGGCGCGCGGAGGG 5259
    ||| ||| ||| |||

RESULT 9
SFU08223       SFU08223 7082 bp DNA BCT 07-SEP-1994
LOCUS          Streptomyces fradiae T59235 tylosin biosynthetic gene region,
DEFINITION     cytochrome P-450, dTDP-glucose synthase, dTDP-glucose dehydratase,
                and thioesterase genes, complete cds.
ACCESSION      U08223
NID            9473596
VERSION        U08223.1 GI:473596
KEYWORDS       Streptomyces fradiae.
SOURCE         Streptomyces fradiae.
ORGANISM       Streptomyces fradiae.
REFERENCE      1. (bases 1 to 7082)
AUTHORS        Merson-Davies, L.A. and Cundliffe, E.
TITLE          Analysis of five tylosin biosynthetic genes from the tyLBA region
               of the Streptomyces fradiae genome
JOURNAL        Mol. Microbiol. 13, 349-355 (1994)
MEDLINE        95075319
REFERENCE      2. (bases 1 to 7082)
AUTHORS        Merson-Davies, L.A.
TITLE          Direct Submission
JOURNAL        Streptomyces fradiae
FEATURES       Location/Qualifiers
               source          1..7082
               /organism="Streptomyces fradiae"
               /strain="T59235"
               /db_xref="taxon:1906"
               944. .2197
               /note="ORF1"
               /codon_start=1
CDS            944. .2197
```

```
/transl_table=11
/product="cytochrome P-450"
/protein_id="AAA21341.1"
/db_xref="PID:g473597"
/db_xref="GI:473597"
/transl_table=11
/product="MTOTLEAKPEPERTANSVHPLAQAGARGILLEWFARARA
EAPVWDSRQAWQVRDDYLTSTNQLQSSDFSPFPPEELAILMGPGTGGID
PPHGLRLKLSQAFTRIATLEPIAEITRGLDGLREKQIDVSDLAYPLPIV
IAELGIPAEORDLREVDVILNNEGMEYNLPDQFSETMGPAIKENGVDLYRRIAL
KRETPDMLGLIEAEVGRRLTDEIVNIVALLTAGHISATLLGNLFLVLDHR
EAOALRADRLIPCAIEETLRYRSPENIIFELAKEDTDILGHKPKACOMVAMTASA
NRDSAHFSDPTDFVRRQPNKHMSEGHGHIHCLGSFLARLEAKVLELFFDFSYRV
EHDEVEFEDELTAARLPVTVIRH"
2677..3843
/note="ORF2, GTG start codon"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAA21342.1"
/db_xref="PID:g473598"
/db_xref="GI:473598"
/transl_table=11
/product="MTGLPRPAVRVPFHDLRDVHAATGVSEIGGALLRVARGRVLL
CAELAAFEERAEVCGNAHCYAVGSLDDARLALMALGVGEGDEVIFSHFIAWLA
VSATGATVPVEGDPGPGFGLDPRLEAALTPTKRAVMPVHLGHVPDLDVPG
AFAPGHVAVVEDAAQYARYRGRIGSGHRTAFYFGKNGALGDGAVVTSDEL
ADRLRLNYGARERYRHEERTNSRLDELAALSVKLPYLDANWTRTAAAYGE
ALAGLPVTPGVRVAEPVHOYVLRSPVDRRLRRLAEAGVETLVHPVAVHAGSAY
AGAGPCPAGGLPRAERLAGEVLSLPIGPHLDEAVEVIAVQSAALDSWBERGP"
3840..4751
/note="ORF3; GTG start codon"
/codon_start=1
/transl_table=11
/product="dUDP-glucose synthase"
/protein_id="AAA21343.1"
/db_xref="PID:g473599"
/db_xref="GI:473599"
/transl_table=11
/product="MNDPRPRAMKGITILAGSGTRLRLPTGLSKQLLPVYDKPMIY
PLSVLMAGRIQIISCKDHLDFSLTGEDRGLGLSTVAEQREPRGIAEAFILGA
RHGGDDAALIGDNVHFGPGFSSVLTGTVARLDGCELFYGVKDAHRYGVEIDSG
RLLSLEPRPRLEGRILVLYTNDVVEIARTISPSARGELEITDVKVLYEQGAA
HGAGAVAWLDMGTHDSLQAGVYQLLEQGERIACIEIETAMRWGFIASRQCYRLG
QELRSSVGSYIIDVAMRGAAADSAQ"
4812..5813
/note="ORF4"
/codon_start=1
/transl_table=11
/product="dUDP-glucose dehydratase"
/protein_id="AAA21344.1"
/db_xref="PID:g473600"
/db_xref="GI:473600"
/transl_table=11
/product="MRVLVTGCGAGFIGSHFTGOLLTGAYPDLGATRTVVLDKLTVAGN
PANLHVAGHPDLEVRGDIADHGWRLMEGVGLVHFAEASHVDRIESEAVRT
NEGVRVLLQAAVDAGVGRFVHISTDEVYSGIAEGSWPEDHPVAPNSPYATKAASDL
LALAYHRTYGLDVRTRCSNNYQPYEKAPELFTNLLDGLPPLYGDGNTREWL
HVDHCGVALVAGGCGPGVIYINIGGTELTNAELTDRIELCGLADRSALRRVADRP
HRRYSVDTTKIREELGYAPRTGITEGLAGTVAWYRDNRANWEPLKRSPPGRELERA"
5875..6636
/note="ORF5, GTG start codon"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAA21345.1"
/db_xref="PID:g473601"
/db_xref="GI:473601"
/transl_table=11
/product="MWLCRYAPVPGTPAHLRVCFPHAGGSARAYRPFPALELAAGVET
HAVQYQGRDRKEPFARLEAERVLPELRLLDAPDGPVALFGHSMGAVVAYET
ARLLHRSAGPRPAGLIGSRAPTADRTETALHGLRELLAEIRLQGTDPGALADEE
VLRVLPALRGDYAVRGYRVHVPGRPCPLTVETGDADPNVTLPEAEAWRELITGAF
ALRVPPGCHYLNDQREAVCTRIETLHRGSKAH"
1008 a 2479 c 2535 g 1060 t

BASE COUNT
ORIGIN

Query Match 8.8%; Score 39.8; DB 1; Length 7082;
Best Local Similarity 50.3%; Pred. No. 7.2;
Matches 98; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 196 GTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTACTTGCCTCCCGGAGGCTTCGCCGTAC 255
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3532 GTGCCAGAGCGCGCGTGCAGGAGCGCGTCTGGCATCAGTACGTCTGCGCAGCCCGTAC 3591
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 256 CGGCTGCACCTTGTGGCTCGCCCTGTGGCTCGTGGGATGTGCGCGCGAGACCATGTCA 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3592 CGGACCGGTTGGCGCGCGCGTGGCGGAGCGGGGTGGAGACCTGTGCTACTATCCG 3651
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 CTGCCAGTCGCGGGGATGACCTGGACCGGCGCTGCTGCTGCTGCGAGCCTCGAGG 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3652 GTGGCGTGCACGCGTACGCGGCGGTACGCGGCGCGGCGCGTCCCGCGGGGCTG 3711
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 TCGCGCGCAGCGCGT 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3712 CGCGCGCGGAGCGT 3726
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
CVCMYC CVCMYC 1320 bp mRNA MAM 07-OCT-1996
LOCUS C.familiaris mRNA for c-myc proto-oncogene.
DEFINITION X95367
ACCESSION g1171522
NID X95367.1 GI:1171522
VERSION c-myc gene; DNA-binding protein.
KEYWORDS dog.
SOURCE Canis familiaris
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1320)
AUTHORS Pirson,I., Coulonval,K., Lamy,F. and Dumont,J.E.
TITLE c-Myc expression is controlled by the mitogenic CAMP-cascade in
thyrocytes
J. Cell. Physiol. 168 (1), 59-70 (1996)
MEDLINE 96257835
REFERENCE 2 (bases 1 to 1320)
AUTHORS Pirson,I.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1996) I. Pirson, IRIH-N-ULB, Blg C- (C4-123),
Campus Erasme, Route de Lennik 808, B- 1070 Brussels- BELGIUM,
BELGIUM
FEATURES
Location/Qualifiers
1..1320
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone_lib="lambda gt11"
/tissue_type="thyroid"
1..1320
/gene="c-myc"
1..1320
/gene="c-myc"
/note="DNA-binding protein"
/codon_start=1
/protein_id="CAA64654.1"
/db_xref="PID:e220184"
/db_xref="PID:g1171523"
/db_xref="GI:1171523"
/db_xref="SWISS-PROT:Q28350"
/transl_table=11
/product="MPLNVSFANRYDLDYDVQVYFYCDEENFYQQQQQSELQPPA
PSEDITKKFELLPTPLSPSRSGLCSPYAVAFSPRGDDGGGSGSTADQLEMY
TELLAGDMVNFQICDPDDETIKNI IQDCWMSGFSAAKLVSEKLAZYQAARKDSG
SPSPARGCGGCTSSLYLQDLASAASECIDPSVVPYPLNDSSSPKPCASPDAAFS
SSDLSLSAESSPRASPEPLAHEETPTTSSDEEEQDEEIDVYVEKRAAPAKR
SSGSPAGCHGKPPHSLVLRKRVSHVTHQHNAAFPSTRTKDTIPAAKRAFLDGRVLK
QISNRKCAKSPRSSSTEENDKRTHNVLQRNELKRSFFALDRDQIPLENNERAPK
VILKKATAYILSVQAEQKLSERKLRKRLKLEQLRNSGA"
300 a 451 c 364 g 205 t

BASE COUNT
ORIGIN
```

```
Query Match      8.8%; Score 39.8; DB 3; Length 1320;
Best Local Similarity 48.18; Pred. No. 8.4;
Matches 113; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 188 GGCCTGGAGTGCAGAGGCTCATCCAGGGGGACCTGGCTACTTGGCCCCGAGGCTTC 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 GACATGGTGAACAGAGCTTCATCTCGACCGGAGGAGACCTTCATCAAAACATC 384

QY 248 GCCGTACCGGTCGACATGTGGCTCCGCTGTGGCTCGTGGATGTGGCGCCGAGACC 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 ATCATCCAGGAGTGATGTGGAGCGGCTCTCGGCGCGCCCAAGCTCTCTCGGAGAG 444

QY 308 ACATGTCACTGCCAGTGGCGGGATGACTGACCGGAGCGGCTGTGTGTGTGCAG 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 CTGGCTCTACCAAGGCTGCGCGCAAGACAGCGCAGCGCCCGCTCGCGGGCCC 504

QY 368 CCCTGAGTTCGGCGCGAGCGCTGCACAGCGGCGGAGGCGGTTCACAGTCCG 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 GCGGCTGCTCACCTCCAGCCTGTACTGTGAGGACCTGAGCGCGCGCCCTCCG 559

RESULT 11
HVU49482      806 bp      mRNA      PLN      10-SEP-1996
LOCUS      Hordeum vulgare low temperature-responsive RNA-binding protein
DEFINITION      (bit801) mRNA, complete cds.
ACCESSION      U49482
NID      g1229137
VERSION      U49482.1 GI:1229137
KEYWORDS      barley.
SOURCE      Hordeum vulgare
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Hordeum.
REFERENCE      1 (bases 1 to 806)
AUTHORS      Dunn, M.A., Brown, K., Lightowler, R. and Hughes, M.A.
TITLE      A low-temperature-responsive gene from barley encodes a protein
with single-stranded nucleic acid-binding activity which is
phosphorylated in vitro
JOURNAL      Plant Mol. Biol. 30 (5), 947-959 (1996)
MEDLINE      96270373
REFERENCE      2 (bases 1 to 806)
AUTHORS      Dunn, M.A. and Hughes, M.A.
TITLE      Direct Submission
JOURNAL      Submitted (20-FEB-1996) M. Allison Dunn, Biochemistry and Genetics,
University of Newcastle upon Tyne, Medical School, Framlington
Place, Newcastle upon Tyne NE2 4HH, UK
FEATURES      Location/Qualifiers
source      1..806
            /organism="Hordeum vulgare"
            /cultivar="Igr1"
            /db_xref="taxon:4513"
            /tissue_type="leaf meristem"
            70..555
            /gene="b1t801"
            /gene="b1t801"
            /note="glycine-rich RNA-binding protein"
            /codon_start=1
            /product="low temperature-responsive RNA-binding protein"
            /protein_id="AA07749.1"
            /db_xref="PID:g1229138"
            /db_xref="GI:1229138"
            /translation="MADVEYRCFVGLRWATDDQSLQNAFSKYGDVIDSKIITDRETG
            RSRGFGVTFASDEAMRAIEAMNQDLDCRNITVNEAQRSSDGGFGGGGGYGG
            QRRREGGGYGGGGGGYGGRRGGGGYGSRRDGGGGYGGGGGGYGGRRGGGGGNWR
            E"
BASE COUNT      121 a      243 c      286 g      156 t
ORIGIN
```

```
Query Match      8.7%; Score 39.4; DB 7; Length 806;
Best Local Similarity 47.7%; Pred. No. 11;
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 172 GGCATAAGCAGCATTGGCTGGAGTGCAGAGCGTCACCTCCAGGGGGACCTGGCTAC 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 GGCCATGACGGCCAGGACCTGGAGCGCGTAACATCACCGTCAACGAGGCCAGTCCCG 323

QY 232 TTGCCCCGAGGCTTCGCGTACCGGCTGACCTTGTGGCTCCGCTGTGGTCTGTGGGA 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 CGGCTCCGAGGAGGCGGCTTCGGCGCGCGCGGCTACGAGGCCAGCGCTCG 383

QY 292 TGTGCGCGCGAGACCATGTCTACCTGCGGCGGCGATGGATGGACCGGAGCGG 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 TGAGGCGGAGGCGGCTACGAGGCGGCGGCTACGAGGTGGCGGCGGCGGCGGCGG 443

QY 352 CTGCTCTGTGTCAGCCTGAGTTCGCGGCGGCGTGCACAGCGCGGCGGAGCGG 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 CGGCGGCGGCTACGCGAGCGTGCACGCGGCGGCGGCGGCTACGCGGCGGCGG 503

QY 412 C 412
Db 504 C 504

RESULT 12
MUSHOXAA      2578 bp      mRNA      ROD      23-MAY-1995
LOCUS      Mus musculus homeobox protein (Hoxa10) mRNA, complete cds.
DEFINITION      L08757
ACCESSION      L08757
NID      9825647
VERSION      L08757.1 GI:825647
KEYWORDS      homeobox protein.
SOURCE      Mus musculus (strain CD-1, sub_species domesticus) adult kidney
            CDNA to mRNA.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 2578)
AUTHORS      Benson, G.V., Nguyen, T.H. and Maas, R.L.
TITLE      The expression pattern of the murine Hoxa-10 gene and the sequence
recognition of its homeodomain reveal specific properties of
Abdominal B-like genes
JOURNAL      Mol. Cell. Biol. 15 (3), 1591-1601 (1995)
MEDLINE      95166244
COMMENT      On May 23, 1995 this sequence version replaced gi:431359.
FEATURES      Location/Qualifiers
source      1..2578
            /organism="Mus musculus"
            /strain="CD-1"
            /sub_species="domesticus"
            /db_xref="taxon:10090"
            /dev_stage="adult"
            /tissue_type="kidney"
            1..985
            /gene="Hoxa10"
            /number=1
            1..56
            /gene="Hoxa10"
            1..2578
            /gene="Hoxa10"
            /product="homeobox protein"
            29..2578
            /gene="Hoxa10"
            57..1256
            /gene="Hoxa10"
            /codon_start=1
            /product="homeobox protein"
            /protein_id="AAA67125.1"
            /db_xref="PID:g567213"
            /db_xref="GI:567213"
            /translation="MCSSESPAANSFLVDSLISGRGEAGVGSGAGGGGYAHGG
            VYLPASDLFYGLQSGFLPAGLSKRNAPSPPGGGGGGGGLPGTHGYAPAPLDLW
```





```
'repeat_region complement(10863..10925)
repeat_region /rpt_family="GC-rich"
repeat_region 11162..11208
repeat_region /rpt_family="LINE2"
repeat_region complement(11339..11368)
repeat_region /rpt_family="GAA(A)n"
repeat_region complement(11369..11666)
repeat_region /rpt_family="AluSx"
repeat_region complement(11972..12274)
repeat_region /rpt_family="AluSq"
repeat_region 12696..12797
repeat_region /rpt_family="MIR"
repeat_region 13059..13138
repeat_region /rpt_family="MIR"
repeat_region 13220..13331
repeat_region /rpt_family="LINE2"
repeat_region complement(13416..13723)
repeat_region /rpt_family="AluY"
repeat_region complement(13730..14018)
repeat_region /rpt_family="AluJo"
repeat_region complement(14074..14250)
repeat_region /rpt_family="AluJo"
repeat_region complement(14254..14551)
repeat_region /rpt_family="AluSq"
repeat_region complement(14566..14867)
repeat_region /rpt_family="AluY"
repeat_region complement(14888..15027)
repeat_region /rpt_family="FRAM"
repeat_region complement(15036..15149)
repeat_region /rpt_family="MIR"
repeat_region 15988..16261
repeat_region /rpt_family="AluJo"
repeat_region 16276..16398
repeat_region /rpt_family="AluSq/x"
repeat_region complement(16583..16879)
repeat_region /rpt_family="AluSq"
repeat_region complement(17259..17285)
repeat_region /rpt_family="(GA)n"
repeat_region complement(17286..17598)
repeat_region /rpt_family="AluJo"
repeat_region complement(17603..18207)
repeat_region /rpt_family="L1"
repeat_region complement(18217..18348)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(18357..18424)
repeat_region /rpt_family="LINE2"
repeat_region complement(19079..19234)
repeat_region /rpt_family="FRAM"
repeat_region complement(19597..19659)
repeat_region /rpt_family="(GGGA)n"
repeat_region complement(19661..19962)
repeat_region /rpt_family="AluSx"
repeat_region 20116..20202
repeat_region /rpt_family="MIR"
repeat_region 20336..20550
repeat_region /note="predicted exon, program: grail2exons_human_1.3,
repeat_region frame: 0, quality: good, score: 51.000"
repeat_region complement(20724..20837)
repeat_region /rpt_family="(GA)n"
repeat_region 20895..21067
repeat_region /rpt_family="AluJo"
repeat_region 21343..21639
repeat_region /rpt_family="AluSq"
repeat_region complement(21725..21831)
repeat_region /rpt_family="MIR"
repeat_region 23022..23156
repeat_region /rpt_family="AluSq/x"
repeat_region 23157..23443
repeat_region /rpt_family="AluSx"
repeat_region complement(23448..23576)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(24313..24455)
repeat_region /note="predicted exon, program: grail2exons_human_1.3,
misc_feature
```

```
frame: 0, quality: excellent, score: 93.000--DDS
similarity to overlapping ESTs:-(24455..24153) AAL21509
zk88c10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA
clone 489906 3' similar to WP: ZK593.7 CE06628 YEAST
JTA107 LIKE; (303..1); 99% identity.--(24455..24154)
AA215299 zrf94b05.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone
IMAGE:683313 5' similar to WP:ZK593.7 CE06628 YEAST JTA107
LIKE; (165..1); 88% identity.--Additional EST
matches:-A1018131, A1004446, AA989671, A1036173"
complement(join(24313..24455,26758..26825,31020..31095,
31194..31344))
/note="Hypothetical human protein most similar to WP:
ZK593.7 CE06628"
/codon_start=1
/evidence=not_experimental
/product="R30783_1"
/protein_id="AAC25622.1"
/db_xref="PID:g3289993"
/db_xref="GI:3289993"
/translation="MSAGARRNOROMFPSNAPRATHAPRMVGEVKGRAVPAKSH
TARQDGEKGEHLGVQVHRQDDPGKVPGRPSGILKGFPLNLVLDGTIEYMRDPDD
QYKLTEDTRQLGLVVCRTGSVVLICPDQDMEAIPIPFITQQDA"
24891..25193
repeat_region /rpt_family="AluY"
repeat_region complement(25347..25649)
repeat_region /rpt_family="AluJb"
repeat_region 25974..26083
repeat_region /rpt_family="AluJo/FLAM"
repeat_region complement(26087..26384)
repeat_region /rpt_family="AluSx"
repeat_region complement(26397..26433)
repeat_region /rpt_family="AT-rich"
misc_feature complement(26708..26825)

Query Match 8.7%; Score 39.2; DB 11; Length 42189;
Best Local Similarity 47.9%; Pred. No. 8.2;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 190 CCTGGAGTCCAGAGCGTCACTCCAGGGGGGACCTGGCTACTTGCCTCGGCTTCGC 249
Db 10975 CCGCGCTCCAGTGTCCCGCACCTGCGCCACGTGGCTCTCACACCTGCGGCCCG 10916
QY 250 CGTCACTCCAGTGTGGCTTCCGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCT 309
Db 10915 CCGCGCGCACGGCGCCAGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10856
QY 310 ATGTCACTCCAGTGTGGCGGGGATGACCTGGACCGGAGCGCGCTGTGTGTGCGAG 369
Db 10855 CGCGGGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10796
QY 370 CTGAGTCCGGCGCAGCGGTGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
Db 10795 GCGCGCGCGCGCGCGCGCGCTGCTGTGCTGCGGCTCGCGCGCGCGCGCGCGCG 10740

Search completed: October 1, 1999, 15:30:36
Job time: 12752 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 10:59:16 ; Search time 14.12 Seconds  
(without alignments)  
181.169 Million cell updates/sec

Title: US-09-099-898-2  
Perfect score: 581  
Sequence: 1 MKALCILLLPVLGLVSSKT.....CHCQAGMDWTGACRCRVQ 108  
Scoring table: BLOSUM62

Searched: 188963-seqs, 23686106 residues

Database: A\_Geneseq36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581	100.0	108	1 W87710	A cysteine rich so
2	581	100.0	109	1 Y12933	Amino acid sequenc
3	318	54.7	114	1 W87707	A cysteine rich so
4	315	54.2	114	1 W87708	A cysteine rich so
5	284.5	49.0	111	1 W87709	A cysteine rich so
6	273.5	47.1	105	1 W87706	A cysteine rich so
7	245.5	42.3	111	1 W87705	A cysteine rich so
8	227	39.1	52	1 Y12613	Human 5' EST seque
9	202.5	34.9	111	1 W87704	A cysteine rich so
10	82	14.1	2321	1 W49698	Human Notch3 prote
11	82	14.1	1872	1 W68510	Partial human Notc
12	76.5	13.2	60	1 R98207	Nucleotide used in
13	73	12.6	131	1 W4444	Mouse Tie receptor
14	72	12.4	1193	1 R91427	Kalinin/laminin 5
15	72	12.4	1111	1 R91428	Kalinin/laminin 5
16	71.5	12.3	519	1 W30826	The novel tyrosina
17	71.5	12.3	2476	1 W67738	Pig p105 zona pell
18	71	12.2	397	1 R27139	ALP of Lysobacter
19	69	11.9	1094	1 R39821	Truncated tie rece
20	68.5	11.8	2471	1 Y06816	Human Notch2 (humN
21	68	11.7	1138	1 R39820	tie receptor kinas
22	68	11.7	1122	1 R73954	Human tie tyrosine
23	68	11.7	102	1 W24566	Serine protease C-
24	67	11.5	1801	1 W50895	Rat laminin B2 cha
25	67	11.5	915	1 Y13350	Amino acid sequenc
26	66.5	11.4	488	1 R67757	Human fetal brain
27	66.5	11.4	524	1 R67758	Human fetal brain
28	66.5	11.4	670	1 R67759	Human fetal brain
29	66.5	11.4	769	1 R75332	Human fetal brain
30	66	11.4	177	1 R40167	Recombinant growth
31	66	11.4	297	1 W20066	Thielavia terrestr
32	65.5	11.3	1080	1 P50296	Mouse epidermal gr
33	65.5	11.3	1964	1 W95557	Mus musculus notch
34	65	11.2	289	1 R52633	Guinea pig PH-30,
35	65	11.2	2707	1 W27161	Mouse receptor ME2
36	65	11.2	1148	1 W87895	Human JAGGED2 prot
37	64.5	11.1	375	1 W44141	Thermotable alkali
38	64.5	11.1	2799	1 W81867	Human tumour suppr
39	64	11.0	688	1 R80361	Tick WGL+ antigen.
40	64	11.0	213	1 R45359	Wheat germ aggluti
41	64	11.0	197	1 W36951	Protein encoded by
42	64	11.0	685	1 W80813	Nucleotide sequenc
43	64	11.0	659	1 W94497	Human delta-2 matu

44 64 11.0 685 1 W94507 Human delta-2 prot  
45 64 11.0 500 1 W94496 Human delta-2 prot

## ALIGNMENTS

RESULT 1  
W87710  
ID W87710 standard; Protein: 108 AA.  
AC W87710;  
DT 09-WAR-1999 (first entry)  
DE A cysteine rich soluble protein designated C23.  
KW Cysteine rich soluble protein; CRSP; C23; cell development;  
KW mammalian immune system; antibody; abnormal proliferation; cancer;  
KW inflammation; degeneration; regeneration; atrophy.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Protein 1..118  
FT Note= "mature protein"

W09858061-Al.  
23-DEC-1998.  
18-JUN-1998: U12236.  
09-OCT-1997; US-061641.  
19-JUN-1997; US-878730.  
19-JUN-1997; US-878878.  
PA (SCHE) SCHERING CORP.  
PI Franz-Bacon K, Gorman DM, McClanahan TK;  
DR WPI: 99-095339/08.  
N-PSDB; V84059.  
New cysteine-rich soluble proteins - used to modulate proliferation,  
differentiation, trafficking and development of cells, e.g. for  
treating inflammation, cancer and degeneration  
Claim 1: Page 18-19; 119pp: English.  
The present sequence represents a cysteine rich soluble protein (CRSP)  
designated C23. CRSP proteins, and their (ant)agonists, are used to  
modulate physiology, differentiation, trafficking and development  
of cells (including those in culture), particularly cells of the  
mammalian immune system. They are used for treatment of abnormal  
proliferation (cancer, inflammation or degeneration), regeneration,  
degeneration and atrophy. The proteins are also used to raise, or  
detect, antibodies, to design oligonucleotides for library screening,  
in drug screens and to isolate cognate receptors. The antibodies are  
used for affinity purification of CRSP, to screen expression libraries,  
to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
to produce anti-idiotypic antibodies (useful for diagnosis), in  
competitive drug screens, and as therapeutic modulators.

Query Match 100.0%; Score 581; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 8.3e-50; Indels 0; Gaps 0;  
Matches 108; Conservative 0; Mismatches 0;  
QY 1 MKALCILLLPVLGLVSSKTLCMEAINERIOEVAGSLIFRAISSIGLECSQVTSRGDL 60  
DB 1 MKALCILLLPVLGLVSSKTLCMEAINERIOEVAGSLIFRAISSIGLECSQVTSRGDL 60  
QY 61 ATCPRGFAVTGCTCGSACGSDVRAETTCQCAGMDWTGACRCRVQ 108  
DB 61 ATCPRGFAVTGCTCGSACGSDVRAETTCQCAGMDWTGACRCRVQ 108

RESULT 2  
Y12933  
ID Y12933 standard; Protein: 109 AA.  
AC Y12933;  
DT 17-JUN-1999 (first entry)  
DE Amino acid sequence of a human secreted peptide.  
KW Human secreted protein; cancer; immune disorder; infection;  
KW inflammatory disorder; skin disorder; tumour; atherosclerosis;  
KW restenosis; autoimmune disorder; Alzheimer's disease;  
KW peripheral neuropathy; trauma; spinal cord injury; allergy;

KW • hematopoietic disorder; skeletal disorder; neurological disorder;  
 KW arthritic disorder; asthma; immunodeficiency disease; AIDS;  
 KW transplant rejection; ss.

OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..23 "signal peptide"  
 FT Protein 24..108  
 FT /note= "secreted protein"

PN WO9911293-A1.  
 PD 11-MAR-1999.  
 PF 03-SEP-1998; U18360.  
 PR 12-SEP-1997; US-058974.  
 PR 05-SEP-1997; US-057626.  
 PR 05-SEP-1997; US-057663.  
 PR 05-SEP-1997; US-057669.  
 PR 12-SEP-1997; US-058666.  
 PR 12-SEP-1997; US-058667.  
 PR 12-SEP-1997; US-058973.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,  
 PI Rosen GA, Ruben SM, Shi Y;  
 DR WPI: 99-204988/17.  
 DR N-PSDB: X31720.  
 PT New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. neurological disorders,  
 PT tumours, immune disorders, inflammation or haematological disorders  
 PS Claim 11; Page 190; 215pp; English.  
 CC Y12914-68 represent human secreted proteins. The polypeptides and  
 CC their corresponding polynucleotides are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g. by protein or gene  
 CC therapy. Pathological conditions can also be diagnosed by determining  
 CC the amount of the new polypeptides in a sample or by determining the  
 CC presence of mutations in the new polynucleotides. Specific uses are  
 CC described for each polynucleotide, based on which tissues they are  
 CC most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, immune disorders, infection,  
 CC inflammatory disorders, skin disorders, tumours, atherosclerosis,  
 CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral  
 CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic  
 CC disorders, skeletal disorders, neurological disorders, arthritic  
 CC disorders, asthma, immunodeficiency diseases, AIDS and transplant  
 CC rejection. The polypeptides are also useful for identifying their  
 CC binding partners.  
 SQ Sequence 109 AA;

Query Match 100.0%; Score 581; DB 1; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-50;  
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKALCLLLPVLGLVSSKTLCSMEAINERIOEVAGSLIFRAISSIGLEQCQVTSRGDL 60  
 DB 1 MKALCLLLPVLGLVSSKTLCSMEAINERIOEVAGSLIFRAISSIGLEQCQVTSRGDL 60  
 QY 61 ATCPRGFVGTCTGSGACGSDVRAETTCCHCCQAGMDWTGACRCRVQ 108  
 DB 61 ATCPRGFVGTCTGSGACGSDVRAETTCCHCCQAGMDWTGACRCRVQ 108

RESULT 3  
 ID W87707  
 AC W87707 standard; Protein; 114 AA.  
 DT 09-MAR-1999 (first entry)  
 DE A cysteine rich soluble protein designated C19.  
 KW Cysteine rich soluble protein; CRSP; C19; cell development;  
 KW mammalian immune system; antibody; abnormal proliferation; cancer;  
 KW inflammation; degeneration; regeneration; degeneration; atrophy.  
 OS Mus sp.  
 FH Key Location/Qualifiers  
 FT Protein 1..20  
 FT /note= "mature protein"

PN WO9858061-A1.  
 PD 23-DEC-1998.  
 PF 18-JUN-1998; U12236.  
 PR 09-OCT-1997; US-061641.  
 PR 19-JUN-1997; US-878730.  
 PR 19-JUN-1997; US-878878.  
 PA (SCHE) SCHERING CORP.  
 PI Franz-Bacon K, Gorman DM, McClanahan TK;  
 DR WPI: 99-095339/08.  
 DR N-PSDB: V84056.  
 PT New cysteine-rich soluble proteins - used to modulate proliferation,  
 PT differentiation, trafficking and development of cells, e.g. for  
 PT treating inflammation, cancer and degeneration  
 PS Claim 1; Page 15; 119pp; English.  
 CC The present sequence represents a cysteine rich soluble protein (CRSP)  
 CC designated C19. CRSP proteins, and their (ant)agonists, are used to  
 CC modulate physiology, differentiation, trafficking and development  
 CC of cells (including those in culture), particularly cells of the  
 CC mammalian immune system. They are used for treatment of abnormal  
 CC proliferation (cancer, inflammation or degeneration), regeneration,  
 CC degeneration and atrophy. The proteins are also used to raise or  
 CC detect antibodies, to design oligonucleotides for library screening,  
 CC in drug screens and to isolate cognate receptors. The antibodies are  
 CC used for affinity purification of CRSP, to screen expression libraries,  
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
 CC competitive drug screens, and as therapeutic modulators.  
 SQ Sequence 114 AA;

Query Match 54.7%; Score 318; DB 1; Length 114;  
 Best Local Similarity 55.8%; Pred. No. 3.4e-24;  
 Matches 58; Conservative 16; Mismatches 30; Indels 0; Gaps 0;  
 QY 4 LCCLLLPVLGLVSSKTLCSMEAINERIOEVAGSLIFRAISSIGLEQCQVTSRGDLATC 63  
 DB 8 LLFLFLLPELLGSSMPLCPIDDAIDKIKQDFNSLPFAIKNIGLNCWTVSSRGKLASC 67  
 QY 64 PRGFVGTCTGSGACGSDVRAETTCCHCCQAGMDWTGACRCRVQ 107  
 DB 68 PEGTAVLSCGSGACGSDVREKVCVCCQARIDWTAARCKLQ 111

RESULT 4  
 ID W87708  
 AC W87708 standard; Protein; 114 AA.  
 DT 09-MAR-1999 (first entry)  
 DE A cysteine rich soluble protein designated C19.  
 KW Cysteine rich soluble protein; CRSP; C19; cell development;  
 KW mammalian immune system; antibody; abnormal proliferation; cancer;  
 KW inflammation; degeneration; regeneration; degeneration; atrophy.  
 OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT Protein 1..17  
 FT /note= "mature protein"

PN WO9858061-A1.  
 PD 23-DEC-1998.  
 PF 18-JUN-1998; U12236.  
 PR 09-OCT-1997; US-061641.  
 PR 19-JUN-1997; US-878730.  
 PR 19-JUN-1997; US-878878.  
 PA (SCHE) SCHERING CORP.  
 PI Franz-Bacon K, Gorman DM, McClanahan TK;  
 DR WPI: 99-095339/08.  
 DR N-PSDB: V84057.  
 PT New cysteine-rich soluble proteins - used to modulate proliferation,  
 PT differentiation, trafficking and development of cells, e.g. for  
 PT treating inflammation, cancer and degeneration  
 PS Claim 1; Page 16; 119pp; English.  
 CC The present sequence represents a cysteine rich soluble protein (CRSP)  
 CC designated C19. CRSP proteins, and their (ant)agonists, are used to  
 CC modulate physiology, differentiation, trafficking and development



CC of cells (including those in culture), particularly cells of the  
 CC mammalian immune system. They are used for treatment of abnormal  
 CC proliferation (cancer, inflammation or degeneration), regeneration,  
 CC degeneration and atrophy. The proteins are also used to raise, or  
 CC detect, antibodies, to design oligonucleotides for library screening,  
 CC in drug screens and to isolate cognate receptors. The antibodies are  
 CC used for affinity purification of CRSP, to screen expression libraries,  
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
 CC competitive drug screens, and as therapeutic modulators.  
 SQ Sequence 114 AA;

Query Match 54.2%; Score 315; DB 1; Length 114;  
 Best Local Similarity 54.1%; Pred. No. 6.6e-24;  
 Matches 60; Conservative 16; Mismatches 31; Indels 4; Gaps 1;

QY 1 MKALCLLLL---PVLGLVSSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTS 56  
 DB 1 MKNLSFLFLFFLVGLGSPMSLCPDDEALSKKINQDFSSLLPAAMKNTVLHCWSVSS 60  
 QY 57 RGLDLPGRFAVGTCTCGSAGSGWDVRAETTCCHQCAGMDWTGARCCRVQ 107  
 DB 61 RGRLASCEGTIVTSCSGSGSGWDVREDTMCCHQCQCSIDWTARCCTLR 111

RESULT 5  
 ID W87709 standard; Protein; 111 AA.  
 AC W87709, 1999 (first entry)  
 DE A cysteine rich soluble protein designated C10.  
 KW Cysteine rich soluble protein; CRSP; C10; cell development;  
 KW mammalian immune system; antibody; abnormal proliferation; cancer;  
 KW inflammation; degeneration; regeneration; atrophy.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT Protein 1..20 /note= "mature protein"  
 FT W09858061-A1.  
 PN 23-DEC-1998.  
 PD 18-JUN-1998; U12236.  
 PF 09-OCT-1997; US-061641.  
 PR 19-JUN-1997; US-878730.  
 PR 19-JUN-1997; US-878878.  
 PA (SCHE ) SCHERING CORP.  
 PI Franz-Bacon K, Gorman DM, McClanahan TK;  
 DR WPI: 99-095339/08.  
 DR N-PSDB: V84058.  
 PT New cysteine-rich soluble proteins - used to modulate proliferation,  
 PT differentiation, trafficking and development of cells, e.g. for  
 PT treating inflammation, cancer and degeneration  
 PS Claim 1; Page 17; 119pp; English.

CC The present sequence represents a cysteine rich soluble protein (CRSP)  
 CC designated C10. CRSP proteins, and their (ant)agonists, are used to  
 CC modulate physiology, differentiation, trafficking and development  
 CC of cells (including those in culture), particularly cells of the  
 CC mammalian immune system. They are used for treatment of abnormal  
 CC proliferation (cancer, inflammation or degeneration), regeneration,  
 CC degeneration and atrophy. The proteins are also used to raise, or  
 CC detect, antibodies, to design oligonucleotides for library screening,  
 CC in drug screens and to isolate cognate receptors. The antibodies are  
 CC used for affinity purification of CRSP, to screen expression libraries,  
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
 CC competitive drug screens, and as therapeutic modulators.  
 SQ Sequence 111 AA;

Query Match 49.0%; Score 284.5; DB 1; Length 111;  
 Best Local Similarity 49.0%; Pred. No. 5.9e-21;  
 Matches 51; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

QY 4 LCLLLPVLGLVSSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTSRGDLAT 62  
 DB 7 LLLIIPLELQINPGSTQCSLSDVMDKKIKDLVLSLEYSPSPISKSLSCASVKSQGRPS 66  
 QY 63 CPRGFVAVTCTCGSAGSGWDVRAETTCCHQCAGMDWTGARCCRV 106  
 DB 67 CPAGNAVTCAGYGGSGWDVLETTCHQCQCSVDVWTARCCHL 110

RESULT 6  
 ID W87706 standard; Protein; 105 AA.  
 AC W87706;  
 DT 09-MAR-1999 (first entry)  
 DE A cysteine rich soluble protein designated C18.  
 KW Cysteine rich soluble protein; CRSP; C18; cell development;  
 KW mammalian immune system; antibody; abnormal proliferation; cancer;  
 KW inflammation; degeneration; regeneration; atrophy.  
 OS Mus sp.

FH Key Location/Qualifiers  
 FT Protein 1..19 /note= "mature protein"  
 FT W09858061-A1.  
 PN 23-DEC-1998.  
 PD 18-JUN-1998; U12236.  
 PF 09-OCT-1997; US-061641.  
 PR 19-JUN-1997; US-878730.  
 PR 19-JUN-1997; US-878878.  
 PA (SCHE ) SCHERING CORP.  
 PI Franz-Bacon K, Gorman DM, McClanahan TK;  
 DR WPI: 99-095339/08.  
 DR N-PSDB: V84055.

PT New cysteine-rich soluble proteins - used to modulate proliferation,  
 PT differentiation, trafficking and development of cells, e.g. for  
 PT treating inflammation, cancer and degeneration  
 PS Claim 1; Page 14; 119pp; English.  
 CC The present sequence represents a cysteine rich soluble protein (CRSP)  
 CC designated C18. CRSP proteins, and their (ant)agonists, are used to  
 CC modulate physiology, differentiation, trafficking and development  
 CC of cells (including those in culture), particularly cells of the  
 CC mammalian immune system. They are used for treatment of abnormal  
 CC proliferation (cancer, inflammation or degeneration), regeneration,  
 CC degeneration and atrophy. The proteins are also used to raise, or  
 CC detect, antibodies, to design oligonucleotides for library screening,  
 CC in drug screens and to isolate cognate receptors. The antibodies are  
 CC used for affinity purification of CRSP, to screen expression libraries,  
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,  
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in  
 CC competitive drug screens, and as therapeutic modulators.  
 SQ Sequence 105 AA;

Query Match 47.1%; Score 273.5; DB 1; Length 105;  
 Best Local Similarity 48.6%; Pred. No. 6.5e-20;  
 Matches 51; Conservative 12; Mismatches 35; Indels 7; Gaps 2;

QY 4 LCLL--LPLVGLVSSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTSRGDLA 61  
 DB 5 LCLFLVSLFLFLVPGNAQCSFSLVDQRIKALSRQEPKTIIS----CTSVSSGRLA 59  
 QY 62 TCPRGFAVGTCTCGSAGSGWDVRAETTCCHQCAGMDWTGARCCRV 106  
 DB 60 SCPAGMVTGTCAGYGGSGWDIRNGNTCHQCQCSVMDWASARCCRM 104

RESULT 7  
 ID W87705 standard; Protein; 111 AA.  
 AC W87705;  
 DT 09-MAR-1999 (first entry)  
 DE A cysteine rich soluble protein designated C2b.  
 KW Cysteine rich soluble protein; CRSP; C2b; cell development;  
 KW mammalian immune system; antibody; abnormal proliferation; cancer;



```
Best Local Similarity 36.3%; Pred. No. 5,6e-13;
Matches 37; Conservative 24; Mismatches 40; Indels 1; Gaps 1;

QY 6 LLLPVLGLVSSKTLCSMEEAINEIRIQF-VAGSLIFRAISGLEQSVTSRGDLATCP 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LICISLLQLWVPVNTDEIIEIVENKVELLANPANYSTVTKTLSTCTSVKTMRWASCP 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 65 RGFAVTGCTCSAGCSWDVRAETTCCHQCAGMDWTGARCVR 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 AGMTATGCACGFACGSGWEIQSGDTCNCLLLVDWTTARCCQL 110.

RESULT 10
W49698
ID W49698 standard; Protein; 2321 AA.
AC W49698;
DT 21-DEC-1998 (first entry)
DE Human Notch3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leucoencephalopathy; therapy.
OS Homo sapiens.
PN FR2751986-Al.
PD 06-FEB-1998.
PR 16-APR-1997; 004680.
PR 01-AUG-1996; FR-009733.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserre E;
DR WPI; 98-133138/13.
DR N-PSDB; V57001.
DR Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leucoencephalopathy
PS Claim 2; Fig 1.1-1.8; 45pp; French.
CC This sequence represents the human Notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of
CC the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC and leucoencephalopathy (CADASIL) type. Blocking expression of a
CC mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC gene or protein can be used to treat CADASIL or related disorders.
SQ Sequence 2321 AA;

Query Match 14.1%; Score 82; DB 1; Length 2321;
Best Local Similarity 29.9%; Pred. No. 8.4;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGCSWDVRAETTC----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1300 VGVPCQQ-TPRGRPCACPPG--LSGPFSCRFPSPGASNASCAAPCLHGGSCRPAPLA 1356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 90 ---HCQCAGMDWTGARC 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 PFFRCACA-QGWTGPRC 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
W68510
ID W68510 standard; Protein; 1872 AA.
AC W68510;
DT 06-JAN-1999 (first entry)
DE Partial human Notch-3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leucoencephalopathy.
OS Homo sapiens.
PN Key Location/Qualifiers
FT Misc_difference 328
FT /note= "encoded by NAN"
```

```
FT Misc_difference 401
FT /note= "encoded by GNN"
FT Misc_difference 403
FT /note= "encoded by GNC"
FT Misc_difference 406
FT /note= "encoded by GNN"
FT Misc_difference 409
FT /note= "encoded by NNT"
FT Misc_difference 420
FT /note= "encoded by GNC"
FT Misc_difference 706
FT /note= "encoded by NNN"
FT Misc_difference 708
FT /note= "encoded by CCN"
FT Misc_difference 719
FT /note= "encoded by CGN"
FT Misc_difference 728
FT /note= "encoded by CNT"
FT Misc_difference 729
FT /note= "encoded by GTN"
FT Misc_difference 759..789
FT /note= "encoded by NNN"
FT Misc_difference 1425
FT /note= "encoded by GNA"
PN FR2751985-Al.
PD 06-FEB-1998.
PF 01-AUG-1996; 009733.
PR 01-AUG-1996; FR-009733.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserre E;
DR WPI; 98-133137/13.
DR N-PSDB; V57163.
DR Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leucoencephalopathy
PS Claim 2; Fig 1a-1g; 42pp; French.
CC This sequence represents a partial human notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of the
CC cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC leucoencephalopathy (CADASIL) type. Blocking expression of a mutated
CC Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC protein can be used to treat CADASIL or related disorders.
SQ Sequence 1872 AA;

Query Match 14.1%; Score 82; DB 1; Length 1872;
Best Local Similarity 29.9%; Pred. No. 6.6;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGCSWDVRAETTC----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1234 VGVPCQQ-TPRGRPCACPPG--LSGPFSCRFPSPGASNASCAAPCLHGGSCRPAPLA 1290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 90 ---HCQCAGMDWTGARC 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1291 PFFRCACA-QGWTGPRC 1306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
R98207
ID R98207 standard; Protein; 60 AA.
AC R98207;
DT 30-DEC-1996 (first entry)
DE Nucleotide used in production of MSH/MoMuLV chimeric sequence.
KW Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW polytropic MX27 provirus; targeted drug delivery; gene therapy;
KW single chain antibody; envelope protein; ss.
OS Synthetic.
PN W09630504-Al.
PD 03-OCT-1996.
```

PF - 22-MAR-1996; U03908.  
PR 24-MAR-1995; US-409648.  
PA (GENE-) GENETIC THERAPY INC.  
PI (UYS-) UNIV SOUTHERN CALIFORNIA.  
PI Anderson W, Chiang YL, Januszski M, Mackrell AJ;  
PI Zhao Y;  
DR WPI: 96-455352/45.  
PR Cell-targeted retroviral vector particles - having envelope protein  
PT modified with targeting polypeptide  
PS Example 2; Page 36; 73pp; English.  
CC Cell targeted retroviral vector particles can be used in gene  
CC therapy to deliver a heterologous gene to a target cell for  
CC expression of a heterologous polypeptide in that cell. The cell  
CC targeted retroviral vector particles comprise an envelope protein  
CC which is modified to contain a targeting polypeptide (a single chain  
CC antibody), or in the case of moloney murine leukaemia virus  
CC (MoMuLV), alpha melanotropin-stimulating hormone (MSH). Two  
CC oligonucleotides (R98207, R98208) were used to substitute sequences in  
CC MoMuLV for MSH sequences. This oligonucleotide was used to replace  
CC residues 574-591 of MoMuLV envelope protein (See W04248).  
SQ Sequence 60 AA;

```

Query Match      13.28; Score 76.5; DB 1; Length 60;
Best Local Similarity 35.19; Pred. No. 0.51;
Matches 20; Conservative 3; Mismatches 19; Indels 15; Gaps 3;

QY 48 GLECSQVTSRGDLATPCPRGFVAVTCCTCGSACGSWDVRAETTCCHCQAGMDWTGARCC 104
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 19 GMACATTTTCG--ATGGSGTAAGGCCCTTAC-----CCCTCGS---TGAAC 60

```

RESULT 13

W74444 standard; Protein; 131 AA.

W74444;

AC 11-MAY-1999 (first entry)

DT Mouse Tie receptor tyrosine kinase fragment.

DE Tie gene; receptor tyrosine kinase; promoter; gene expression;

KW human gene therapy; growth factor.

OS Mus sp.

PN US5877020-A.

PD 02-MAR-1999.

PF 31-MAY-1996; 650598.

PR 31-MAY-1996; US-650598.

PR 22-SEP-1994; US-310717.

PA (UYHE-) UNIV HELSINKI LICENSING LTD.

PI Alitalo K;

DR WPI: 99-189653/16.

PT Tie receptor tyrosine kinase promoter - for directing expression of

PT recombinant DNA in endothelial cells

PS Example 1; Column 6; 26pp; English.

CC This sequence represents a fragment of the mouse Tie receptor tyrosine

CC kinase. The invention relates to Tie receptor tyrosine kinase promoters.

CC The promoters are useful for directing expression of recombinant DNA

CC sequences in endothelial cells. The promoters are useful for production

CC of proteins and peptides which act as anticoagulants, vasodilator

CC inhibitors of thrombosis or restenosis into endothelial cells, blood and

CC tissues. The promoters are useful for directing expression of proteins

CC and peptides for human gene therapy, antigens and markers for endothelial

CC cell tagging, and antisense RNA constructs for use in endothelial cells

CC in vivo and in vitro. The promoters, and vectors and host cells

CC containing them, are useful in gene therapy for promoting expression of

CC various growth factors or receptors or their domains.

CC Sequence 131 AA;

SQ

```
Query Match      12.6%   Score 73; DB 1; Length 131;
Best Local Similarity 28.8%   Pred. No. 2.6;
Matches 23; Conservative 4; Mismatches 19; Indels 34; Gaps 5;
```

```

Db 19 DGCVCPPGF--TGTFCEQACREGFRGSCQCEQCFGTAGCRGLTFCLPDPYGCSC--GSGWR 73
Oy 100 GARC-----CRVQ 107
   ||:|         ||:|
Db 76 GSQCQEACAPDHFAGDCRLQ 95

RESULT 14
R91427
ID R91427 standard; Protein; 1193 AA.
AC R91427;
DT 13-NOV-1996 (first entry)
DE Kalinin/laminin 5 gamma-2 chain.
KW kalinin; laminin; epidermolysis bullosa; junctional; probe;
KR detection; inhibit; monitor; malignancy.
OS Homo sapiens.
PN W09610646-A1.
PD 11-APR-1996.
PF 04-OCT-1995; E03918.
PR 04-OCT-1994; US-317450.
PA (TRYG/) TRYGGVASON K.
PI Kallunki P, Pyke C, Tryggvason K;
PI WPI: 96-209366/21.
DR N-PSDB: TL323.
DR Detection of kalinin or laminin 5 expression in cells - useful to
PT detect, monitor and inhibit the invasive growth of cell in tissue,
PT partic. malignant tissue
PT Disclosure; Fig 4A; 37pp; English.
PS The present sequence is the kalinin/laminin 5 gamma-2 chain. The gamma-
CC chain is of importance to patients suffering from epidermolysis bullo
CC esp. the junctional form (JEB). Probes and antisense gamma-2 sequence
CC derived from this sequence can be used to detect, monitor and inhibi
CC the invasive growth of cells in tissue, partic. malignant tissue.
CC sequence 1193 AA;
SQ

```

```

Query Match      12.4%; Score 72; DB 1; Length 1193;
Best Local Similarity 25.4%; Pred. No. 38;
Matches 31; Conservative 14; Mismatches 47; Indels 30; Gaps 8;

QY    4 LCLLLLPVLGLLVSSKTLG-----SMEEAINERIQEVAGSLIFRAI-----SSIGLECSQSV 54
      ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    10 LCFSLLPAAARATSRREYCDNGKSROCIFDREHROTGN-GFRLCNCNDMTDGIHCEK- 68
      ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY    55 TSRGDLATCPRGF-----AVTGCTGSAGSGMDVRAETTCQCQ-AGMDWTGARCCRV 106
      ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    58 -----CKNGFYHRERDRCLFCNCSNK-GSLSARCDNSGRCSCKPGV--TGAARDRC 116
      ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

QY    107 QP 108
      |
Db    117 LP 118

RESULT 15
R91428
ID R91428 standard; Protein; l111 AA.
AC R91428:
DT 13-NOV-1996 (first entry)
DE Kalinin/laminin 5 gamma-2 chain (alternative form).
KW kalinin; laminin; epidermolysis bullosa; junctional; probe;
KW detection; inhibit; monitor; malignancy.
OS Homo sapiens.
PN W09610646-AL.
PS
ED 11-APR-1996.
PF 04-OCT-1995; E03918.
PR 04-OCT-1994; US-317450.
PA (TRYG/) TRYGGVASON K.
PI Kallunki P., Pyke C., Tryggvason K;
DR WPI: 96-209366/21.
DR N-FSDb: TL3323.
PT Detection of kalinin or laminin 5 expression in cells - useful to
PT detect, monitor and inhibit the invasive growth of cell in tissue,
PT partic. malignant tissue
```

PS Disclosure; Fig 4B; 37pp; English.  
CC The present sequence is an alternative form of kallinin/laminin 5 gamma-2  
CC chain (see R91427). The gamma-2 chain is of importance to patients  
CC suffering from epidermolysis bullosa, esp. the junctional form (JEB).  
CC CC Probes and antisense gamma-2 sequences derived from this sequence can be  
CC used to detect, monitor and inhibit the invasive growth of cells in  
CC tissue, partic. malignant tissue.  
SQ Sequence illl AA;  
SQ Sequence illl AA;

Query Match	12.4%;	Score 72;	DB 1;	Length 1111;
Best Local Similarity	25.4%;	Pred.No.35;		
Matches 31; Conservative	14;	Mismatches 47;	Indels 30;	Gaps 8;

  

Qy	4	LCLLLLVIGLLVSSKTLG-----SMEPAINERIOEVAGSLIERAI-----SSIGLEQOSV	54
		:     :   :   :   :   :   :   :   :	
Dd	10	LCFSLLLPAAATSRREVCDNGKRSRCIFREHLRQTGN--GFRLCNCNDNTDGIHCEK-	68
		:     :   :   :   :   :   :   :   :	
Qy	55	TSRGDLATCPRGF-----AVTGCTGSACGSWDVRAETTCQC-AGMDWTGARCCRV	106
		:     :   :   :   :   :   :   :   :	
Dd	68	-----CKNGFYHRREDRCLPCNCSNK-GSLSARCDNSGRCSCKPGV--TGA CDRC	116
		:     :   :   :   :   :   :   :   :	
Qy	107	QP 108	
Dd	117	Lp 118	

Search completed: October 2, 1999, 11:02:33  
Job time: 197 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 00:40:20 ; Search time 14.11 Seconds  
(without alignments)  
471.064 Million cell updates/sec

Title: US-09-099-898-2

Perfect score: 581

Sequence: 1 MKALCLLLPVLGLVSSKT.....CHCQCAGMDWTGARCCRVQP 108

Scoring table: BLOSUM62

Searched: 201082-seqs, 61543640 residues

Database :

```
Database ;
SF INEMBLY_10;
1: sp archea:*
```

```

1: sp_bacteria:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organella:*
9: sp_phase:*
10: sp_plant:*
11: sp_protein:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	86	14.8	1059	5	P90884
2	84.5	14.5	1476	13	Q90285
3	82.5	14.2	1574	11	O88281
4	82.5	14.2	762	13	O42373
5	78	13.4	1106	5	Q17434
6	76.5	13.2	589	11	O88671
7	76	13.1	2447	13	Q13149
8	75.5	13.0	153	4	O75095
9	75.5	13.0	387	11	O60607
10	74.5	12.8	1698	5	Q94438
11	74.5	12.8	75	5	O96388
12	74.5	12.8	585	11	O35675
13	74.5	12.8	592	11	O88516
14	73.5	12.7	830	4	O43701
15	73.5	12.7	830	4	Q14162
16	73.5	12.7	289	5	P91237
17	71.5	12.3	378	3	O59925
18	71.5	12.3	237	4	O75767
19	71.5	12.3	105	5	Q17187
20	71.5	12.3	1647	5	O45000
21	71.5	12.3	2704	5	O97458
22	71	12.2	341	5	O25032
23	71	12.2	752	13	O42374
24	70.5	12.1	626	10	O49438
25	70	12.0	582	2	O92892
26	70	12.0	709	5	O97444
27	70	12.0	1095	11	O60784
28	70	12.0	3857	11	O88840
29	70	12.0	127	13	O95GK3
30	70	12.0	127	13	O95GK3
31	70	12.0	127	13	O95GK3
32	70	12.0	127	13	O95GK3
33	70	12.0	127	13	O95GK3
34	70	12.0	127	13	O95GK3
35	70	12.0	127	13	O95GK3
36	70	12.0	127	13	O95GK3
37	70	12.0	127	13	O95GK3
38	70	12.0	127	13	O95GK3
39	70	12.0	127	13	O95GK3
40	70	12.0	127	13	O95GK3
41	70	12.0	127	13	O95GK3
42	70	12.0	127	13	O95GK3
43	70	12.0	127	13	O95GK3
44	70	12.0	127	13	O95GK3
45	70	12.0	127	13	O95GK3
46	70	12.0	127	13	O95GK3
47	70	12.0	127	13	O95GK3
48	70	12.0	127	13	O95GK3
49	70	12.0	127	13	O95GK3
50	70	12.0	127	13	O95GK3
51	70	12.0	127	13	O95GK3
52	70	12.0	127	13	O95GK3
53	70	12.0	127	13	O95GK3
54	70	12.0	127	13	O95GK3
55	70	12.0	127	13	O95GK3
56	70	12.0	127	13	O95GK3
57	70	12.0	127	13	O95GK3
58	70	12.0	127	13	O95GK3
59	70	12.0	127	13	O95GK3
60	70	12.0	127	13	O95GK3
61	70	12.0	127	13	O95GK3
62	70	12.0	127	13	O95GK3
63	70	12.0	127	13	O95GK3
64	70	12.0	127	13	O95GK3
65	70	12.0	127	13	O95GK3
66	70	12.0	127	13	O95GK3
67	70	12.0	127	13	O95GK3
68	70	12.0	127	13	O95GK3
69	70	12.0	127	13	O95GK3
70	70	12.0	127	13	O95GK3
71	70	12.0	127	13	O95GK3
72	70	12.0	127	13	O95GK3
73	70	12.0	127	13	O95GK3
74	70	12.0	127	13	O95GK3
75	70	12.0	127	13	O95GK3
76	70	12.0	127	13	O95GK3
77	70	12.0	127	13	O95GK3
78	70	12.0	127	13	O95GK3
79	70	12.0	127	13	O95GK3
80	70	12.0	127	13	O95GK3
81	70	12.0	127	13	O95GK3
82	70	12.0	127	13	O95GK3
83	70	12.0	127	13	O95GK3
84	70	12.0	127	13	O95GK3
85	70	12.0	127	13	O95GK3
86	70	12.0	127	13	O95GK3
87	70	12.0	127	13	O95GK3
88	70	12.0	127	13	O95GK3
89	70	12.0	127	13	O95GK3
90	70	12.0	127	13	O95GK3
91	70	12.0	127	13	O95GK3
92	70	12.0	127	13	O95GK3
93	70	12.0	127	13	O95GK3
94	70	12.0	127	13	O95GK3
95	70	12.0	127	13	O95GK3
96	70	12.0	127	13	O95GK3
97	70	12.0	127	13	O95GK3
98	70	12.0	127	13	O95GK3
99	70	12.0	127	13	O95GK3
100	70	12.0	127	13	O95GK3

Q95071 homo sapien  
Q44446 chironomus  
Q08126 ovls aries  
Q3680 persca amer  
Q73443 homo sapien  
Q94700 paramesciu  
Q5428 arabidopsis  
Q33348 rattus norv  
Q35883 rattus norv  
Q4179 caenorhabdi  
Q61699 brachiosteo  
Q33720 rattus norv  
Q63721 rattus norv  
Q5352 streptomyces  
Q62554 mytilus edu  
Q01610 caenorhabdi

## ALIGNMENTS

RESULT	I
P90884	PRELIMINARY; PRT: 1059 AA.
ID	P90884
AC	P90884;
DT	01-AUG-1998 (TREMBLrel. 07, Created)
DT	01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE	F53B6.2 PROTEIN.
GN	F53B6.2.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Caen
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caen
RC	[1]
RA	WHITE S.;
RL	Submitted (OCT-1996) to the EMBL/GenBank/DDBB databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE; 94150718.
RA	WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M,
RA	BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J.,
RA	CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., J
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE
RA	LIGHTNING J., LLOYD C., MCURRAY A., MORTMORE B., O'CAL
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D.,
RA	SALMOND N., SMITH A., SONNHAMMER E., STADEN R., SULSTON
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATER
RA	WATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOHLDMAN P
RT	"2.2 Mb of contiguous nucleotide sequence from chromosom
RL	elegans";
RT	Nature 368:32-38(1994).
DR	EMBL; Z81086; CAB03121.1; --
DR	PFAM; PF00047; 1g; 1.
SQ	SEQUENCE 1059 AA; 117768 MW; 4CDD1913 CRC32;

```

Query Match      14.8%; Score 86; DB 5; Length 1059;
Best Local Similarity 27.2%; Pred. No. 0.18;
Matches 31; Conservative 12; Mismatches 33; Indels 38; Gaps 9;

QY  20  TLCSMEEAINE-----RIQEVAGSLIFRAISIGILECQSVTSRGDLATCPRGFA-VTG 71
      ||| ||| : : : ||| ||| ||| ||| : : |
Db   80  TICGGEIVSRGQEVVCRSLRTGANFLWRV--DDGTPCQAATSR---AVCSKSCQOIVG 134

QY  72  C-----TCGSACGS-----WDVRAE-TTCHCOCAG-MDWTGA 101
      ||| ||| ||| ||| ||| ||| ||| |||
Db   135 CDGLISSSFREDAG-VCGRGRTCDNGKFIWKVSEYETACASNCDDIVDSGA 187

RESULT      2
090285

```

ID Q90285 PRELIMINARY; PRT; 1476 AA.  
AC Q90285; Q98847;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE PUTATIVE EXTRACELLULAR AND CYTOPLASMIC FRAGMENT OF NOTCH-3 HOMOLOGY (FRAGMENT).  
DE Carassius auratus (Goldfish).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
RN Cyprinidae; Cyprinidae; Cyprinidae; Carassius.  
RP SEQUENCE FROM N.A.  
RC TISSUE-RETINA (10D POST-OPTIC NERVE CRUSH);  
RA SULLIVAN S.A., BARTHEL L.K., LARGENT B.L., RAYMOND P.A.;  
RT "A goldfish Notch-3 homologue is expressed in neurogenic regions of embryonic, postembryonic, and regenerating CNS.";  
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U09191; AAB17010.1; -;  
DR PFAM; PF00023; ank; 6.  
DR PFAM; PF00008; EGF; 11.  
DR PFAM; PF00066; notch; 3.  
KW ANK repeat; Glycoprotein.  
FT NON\_TER 1 1476 1476  
FT SEQUENCE 1476 AA; 160385 MW; D6077129 CRC32;  
SQ SEQUENCE 1476 AA; 160385 MW; D6077129 CRC32;  
  
Query Match 14.5%; Score 84.5; DB 13; Length 1476;  
Best Local Similarity 33.3%; Pred. No. 0.38; Mismatches 32; Indels 17; Gaps 5;  
Matches 28; Conservative 7;  
  
QY 36 AGSLIFRAISSIGLECO-----SVTSRGDLATCPRGFAVTGCT-----CGS-ACGSWDV 83  
DB 357 AGANCERSMNCLEPCYNGSGSDTSRGARCTCIQGGFGPQOHRSGEGSSAPCSNGGL 416  
  
QY 84 RAETTC-----HCOCAGMDWTGARC 103  
DB 417 CTEETSYPFFHCOCIS-GWKGIRC 439  
  
RESULT 3  
ID O88281 PRELIMINARY; PRT; 1574 AA.  
AC O88281;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE MEGF6.  
DE MEGF6.  
GN MEGF6.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;  
RX MEDLINE; 98360089.  
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;  
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";  
RL Genomics 51:27-34(1998).  
DR EMBL; AB011532; BAA32462.1; -;  
DR PFAM; PF00008; EGF; 25.  
DR PROSITE; PS00010; ASX-HYDROXYL; 5.  
DR PROSITE; PS01187; EGF\_CA; 5.  
KW Glycoprotein; EGF-like domain.  
SQ SEQUENCE 1574 AA; 165445 MW; C49E6EA6 CRC32;  
  
Query Match 14.2%; Score 82.5; DB 11; Length 1574;  
Best Local Similarity 35.6%; Pred. No. 0.69; Mismatches 26; Indels 9; Gaps 2;  
Matches 21; Conservative 3;

QY 47 IGLECOVSRTSGDLATCPRGFAVTGCTCGSACGSWDVRAETTCACAGMDWTGARC 105  
DB 813 VGSRCOD-----TCSAGWGTGCOIRCACANDGCHDPTTGRCSCA-PGWTLSCOR 862  
  
RESULT 4  
ID O42373 PRELIMINARY; PRT; 762 AA.  
AC O42373;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE NOTCH RECEPTOR PROTEIN (FRAGMENT).  
GN NOTCH5.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Cyprinidae; Rasborinae; Danio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WESTIN J., LARDELLI M.;  
RL Dev. Genes Evol. 207:51-63(1997).  
DR EMBL; Y10353; CAA71379.1; -;  
DR PFAM; PF00023; ank; 3.  
DR PFAM; PF00008; EGF; 5.  
DR PFAM; PF00066; notch; 3.  
DR 2FIN; ZDB-GENE-980526-78; notch5.  
KW Glycoprotein.  
FT NON\_TER 1 762 762  
FT SEQUENCE 762 AA; 83560 MW; 9FE1F01E CRC32;  
SQ SEQUENCE 762 AA; 83560 MW; 9FE1F01E CRC32;  
  
Query Match 14.2%; Score 82.5; DB 13; Length 762;  
Best Local Similarity 32.1%; Pred. No. 0.32; Mismatches 27; Conservative 9; Indels 17; Gaps 5;  
Matches 27; Conservative 9;  
  
QY 36 AGSLIFRAISSIGLECO-----SVTSRGDLATCPRGFAVTGCT-----CGS-ACGSWDV 83  
DB 142 AGNCERSMNCLEPCYNGSGSCLTTTGARCTCIQGGFGPQOHRSGEGSSAPCSNGGL 201  
  
QY 84 RAETTC-----HCOCAGMDWTGARC 103  
DB 202-CTEETSYPFFHCOCIS-GWKGIRC 224  
  
RESULT 5  
ID Q17494 PRELIMINARY; PRT; 1106 AA.  
AC Q17494;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE B0393.5 PROTEIN.  
GN B0393.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SULSTON J.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

```

RN      [1]
RP      SEQUENCE FROM N.A.
RA      NAKAMURA T., TROWSDALE J.;
RL      Submitted (JUN-1997) to the EMBL/GenBank/DBDJ databases.
RR      EMBL; AB004829; BAA20535.1; -
DR      PFAM; PF00023; ank; 6.
DR      PFAM; PF00008; EGF; 35.
DR      PFAM; PF00066; notch; 3.
DR      PROSITE; PS001010; ASX_HYDROXYL; 22.
DR      PROSITE; PS01187; EGF_CA; 20.
KW      Glycoprotein; EGF-like domain.
FT      NON_TER 1
SQ      SEQUENCE 2447 AA; 262542 MW; 3CDA4F7A CRC32;

Query Match          13.1%; Score 76; DB 13; Length 2447;
Best Local Similarity 26.7%; Pred.No. 5.9; Gaps 6;
Matches 24; Conservative 2; Mismatches 20; Indels 44; Gaps

QY    50 ECQSVTSRGDLATCPRGFAVTC-----GSAGC-----SW 81
      || |   || |   |||
Db     982 ECDSPCKNG-GTCTDGLGYRCTCPAGYNQCNYNLCRVRCHNGSCSHTGATSW 1040
      || |   || |   |||

QY    82 DVRAETTCQCACAGMDWTGARC-----CR 105
      |||| | ||| |
Db     1041 -----TCHCT---MGWTPGYCDVPDMSCR 1061
      |||| | ||| |

RESULT 8
O75095 PRELIMINARY; PRT; 153 AA.
ID O75095 AC
O75095 AC
DT DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DE MEGF6 (FRAGMENT).
GN GN MEGF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN      SEQUENCE FROM N.A.
RP      TISSUE=BRAIN;
RC      MEDLINE; 98360089.
RX RX NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RA "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
RR EMBL; AB011539; BAA32467.1; -.
DR PFAM; PF00008; EGF; 2.
FT NON_TER 1
SQ SEQUENCE 153 AA; 15579 MW; 1AC64E20 CRC32;

Query Match          13.0%; Score 75.5; DB 4; Length 153;
Best Local Similarity 35.7%; Pred.No. 0.34;
Matches 20; Conservative 1; Mismatches 26; Indels 9; Gaps 3;

QY    48 GLECSVTSRGDLATCPRGFAVTCGSGSWSVDVRAETTCQCACAGMDWTGARC 103
      || |   || |   ||| | ||| |
Db     51 GAPCDPVTG----LCLCPPG--RSGATCNLCRRGQFGPSCTLHCDCGG----GADC 97
      || |   || |   ||| | ||| |

RESULT 9
Q06007 PRELIMINARY; PRT; 387 AA.
ID Q06007 AC
Q06007 AC
DT DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE DE NOTCH PROTEIN HOMOLOG 1 (MOTCH A PROTEIN) (FRAGMENT).
FT NOTCH1 OR MOTCH A.
OS Mus musculus (Mouse).
```



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
-RP SEQUENCE FROM N.A.  
RC STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;  
RX MEDLINE: 93178563.  
RA LARDELLI M., LENDAHU U.;  
RT "Moth A and moth B--two mouse Notch homologues coexpressed in a  
wide variety of tissues."  
RL Exp. Cell Res. 204:364-372(1993).  
DR EMBL: X68278; CAA48339.1; -  
DR MGD: MGI:97363; NOTCH1.  
DR PFAM: PF00008; EGF; 6.  
DR PFAM: PF00066; notch; 3.  
KW Differentiation; Neurogenesis; Repeat.  
FT NON\_TER 1 387  
FT NON\_TER 387 387  
SQ SEQUENCE 387 AA; 41497 MW; D1FD6C00 CRC32;  
  
Query Match 13.0%; Score 75.5; DB 11; Length 387;  
Best Local Similarity 29.1%; Pred. No. 0.93; Mismatches 4; Indels 33; Gaps 5;  
Matches 25; Conservative 4; Mismatches 4; Indels 33; Gaps 5;  
  
Qy 48 GLECOV-----TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84  
Db 141 GRCEVINGCRGKCKNGVCVASANTARGFCRCAGF--EGATCENDARTCGSLRCL 198  
Qy 85 AETTC-----HCOCAGMDWTGARC 103  
Db 199 NGTCTGSPRPTCLCLG-SFTGPEC 223  
  
RESULT 10  
Q94438  
ID Q94438 PRELIMINARY; PRT; 1698 AA.  
AC Q94438;  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)  
DE 185 KDA SILK PROTEIN.  
GN SPI85.  
OS Chironomus pallidivittatus (Midge).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Diptera; Nematocera; Chironomoidea; Chironomidae;  
OC Chironominae; Chironomus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SPINAL GLAND;  
RA CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U54640; AAA99803.1; -  
SQ SEQUENCE 1698 AA; 186164 MW; 85E8E520 CRC32;  
  
Query Match 12.8%; Score 74.5; DB 5; Length 1698;  
Best Local Similarity 26.4%; Pred. No. 5.9; Mismatches 25; Indels 33; Gaps 5;  
Matches 23; Conservative 6; Mismatches 25; Indels 33; Gaps 5;  
  
Qy 52 QSVTSRGDLATCPRGFAVTGCTC---GSAC-----GSMVDVRAETTCCHCOC--- 94  
Db 1355 QTNWQTCOCPCASGCTCTGAQVWCSKACKVCVPAQKCKDSPTWD---ESSCSCOCAPN 1411  
Qy 94 -----AGMDWTGARC---CRVOP 108  
Db 1412 MRPPKGCNAGRTWDDATCSEKCAVP 1438  
  
RESULT 11  
Q96388  
-ID Q96388 PRELIMINARY; PRT; 75 AA.  
AC Q96388;

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
DE METALLOTHIONEIN.  
OS Perna viridis.  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
OC Mytilidae; Perna.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHOO H.W., WONG Y.W., PATEL K.H.;  
RT "Green mussel (Perna viridis) metallothionein cDNA";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF036904; AAD02054.1; -  
SQ SEQUENCE 75 AA; 7415 MW; 942586D8 CRC32;  
  
Query Match 12.8%; Score 74.5; DB 5; Length 75;  
Best Local Similarity 27.9%; Pred. No. 0.21; Mismatches 17; Indels 17; Gaps 4;  
Matches 17; Conservative 9; Mismatches 18; Indels 17; Gaps 4;  
  
Qy 51 COSVTSRGDLATCPRGFAVTGCTCGSACGSWDVRAETTCCHCOCAGMDWTGARC---CRVQ 107  
Db 5 CNCIETQ--VCICGTGCGEGRCGDAC-----KCSGCGCGCGSG-----CKVVKCQ 50  
Qy 108 P 108  
Db 51 P 51  
  
RESULT 12  
O35675  
ID O35675 PRELIMINARY; PRT; 585 AA.  
AC O35675;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
DE M-DELTA-LIKE 3 GENE PRECURSOR.  
GN DLL3 OR M-DELTA-LIKE 3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;  
RX MEDLINE: 97417575.  
RA DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;  
RT "Mouse Dll3: a novel divergent Delta gene which may complement the  
function of other Delta homologues during early pattern formation in  
the mouse embryo."  
RL Development 124:3065-3076(1997).  
DR EMBL: Y11895; CAA72637.1; -  
DR MGD: MGI:1096877; DLL3.  
DR PFAM: PF00008; EGF; 6.  
KW Signal; Glycoprotein.  
FT SIGNAL 1 32  
FT SIGNAL 32  
SQ SEQUENCE 585 AA; 61129 MW; 0F5E6187 CRC32;  
  
Query Match 12.8%; Score 74.5; DB 11; Length 585;  
Best Local Similarity 28.8%; Pred. No. 1.9; Mismatches 34; Indels 43; Gaps 8;  
Matches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;  
  
Qy 21 LC-----SMEAINERTQEA--GSLIFRAISSIGLSCQ-----SVTSRGDLATCPRGF-- 68  
Db 245 LCTVPVSTSCSLNSRVPGPASTGCLLPGPCDNGCANGSGSCSTGSGFEACAPRGFYG 304  
Qy 68 ---AVTGCTC-----GSAC-GSWDVRAETTCCHCOCAGMDWTGARC-----CRVOP 108  
Db 305 LRCEVSGVTCADGCPFGNGLCVGGEDPSYVCHCP---PFGQSGNCEKRVDRCLQP 359  
  
RESULT 13  
O88516  
ID O88516 PRELIMINARY; PRT; 592 AA.

```

O88516;
AC 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, last annotation update)
DE DELTA-LIKE 3 ALTERNATE SPLICED FORM 2.
GN DLL3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE: 98324780.
RA KUSUMI K., SUN E.S., KERREBROCK A.W., BRONSON R.T., CHI D.-C.,
RA BULOTSKI M.S., SPENCER J.B., BIRREN B.W., FRANKEL W.N., LANDER E.S.;
RT "The mouse pudy mutation disrupts delta homologue D113 and
RT initiation of early somite boundaries."
RL Nat. Genet. 19:274-278(1998).
DR EMBL: AF068865; AAC40170.1; -.
DR PFAM: PF00008; EGF; 6.
KW Glycoprotein.
SQ SEQUENCE 592 AA; 62069 MW; C1829CC9 CRC32;

Query Match 12.8%; Score 74.5; DB 11; Length 592;
Best Local Similarity 28.8%; Pred. No. 1.9;
Matches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;

QY 21 LC-----SMEERAINRIQVAA--GSLIFRAISSIGLEQC-----SVTSRGDLATCPRGF-- 68
DB 245 LCTVPVSTSSCLNSRVPQPGASTGCLLPQPGCDGNPCANGSGCSETSGFECACPRGFVG 304
QY 68 ---AVTGTC-----GSAC--GSHDVRAETTCCHQCAGMDWTGARC-----CRVQP 108
DB 305 LRCEVSGVTGADGCPFCFNGGLCVGGEDPDPSAYVCHCP---PGFQGSNCEKRVDRCSLQP 359

RESULT 14
O43701
ID O43701 PRELIMINARY; PRT; 830 AA.
AC O43701;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, last annotation update)
DE ACETYL LDL RECEPTOR PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98058897.
RX ADACHI H., TSUJIMOTO M., ARAI H., INOUE K.;
RT "Expression cloning of a novel scavenger receptor from human
RT endothelial cells."
RL J. Biol. Chem. 272:31217-31220(1997).
DR EMBL: D86864; BAA24070.1; -.
KW Signal; Glycoprotein.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 830 ACETYL LDL RECEPTOR.
SQ SEQUENCE 830 AA; 87430 MW; 936CC3DD CRC32;

Query Match 12.7%; Score 73.5; DB 4; Length 830;
Best Local Similarity 33.3%; Pred. No. 3.5;
Matches 26; Conservative 1; Mismatches 28; Indels 23; Gaps 7;

QY 51 CQ--SVTSRGDLATCPRGFAVTGCTCGSACGS--W--DVRAETTC-----CQC 93
DB 63 CQKDEVCKVKGCLCKCKPGF--FGAHCSSRCPCQYWGPDRCSCPCPHGQCEPATGACQC 120
QY 94 AGMDWTGARC---CRVQP 108
DB 121 QADRW-GARCEFFPCACGP 137

O88516;
AC 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, last annotation update)
DE DELTA-LIKE 3 ALTERNATE SPLICED FORM 2.
GN DLL3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE: 98324780.
RA KUSUMI K., SUN E.S., KERREBROCK A.W., BRONSON R.T., CHI D.-C.,
RA BULOTSKI M.S., SPENCER J.B., BIRREN B.W., FRANKEL W.N., LANDER E.S.;
RT "The mouse pudy mutation disrupts delta homologue D113 and
RT initiation of early somite boundaries."
RL Nat. Genet. 19:274-278(1998).
DR EMBL: AF068865; AAC40170.1; -.
DR PFAM: PF00008; EGF; 6.
KW Glycoprotein.
SQ SEQUENCE 592 AA; 62069 MW; C1829CC9 CRC32;

Query Match 12.7%; Score 73.5; DB 4; Length 830;
Best Local Similarity 33.3%; Pred. No. 3.5;
Matches 26; Conservative 1; Mismatches 28; Indels 23; Gaps 7;

QY 51 CQ--SVTSRGDLATCPRGFAVTGCTCGSACGS--W--DVRAETTC-----CQC 93
DB 63 CQKDEVCKVKGCLCKCKPGF--FGAHCSSRCPCQYWGPDRCSCPCPHGQCEPATGACQC 120
QY 94 AGMDWTGARC---CRVQP 108
DB 121 QADRW-GARCEFFPCACGP 137

```

Search completed: October 2, 1999, 02:58:58  
Job time: 8318 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 11:02:37 ; Search time 9.08 Seconds  
(without alignments)  
336.231 Million cell updates/sec

Title: US-09-099-898-2  
Perfect score: 581  
Sequence: 1 MKALCLLLPLVGLLVSSKT.....CHCQAGMDWTGRCRQVP 108  
Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database: SwissProt27\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	86	14.8	934	1 CO6_HUMAN	P13671 homo sapien
2	80.5	13.9	456	1 PRTC_BOVIN	P00745 bos taurus
3	75.5	13.0	2139	1 CRB_DROME	P10040 drosophila
4	75.5	13.0	2531	1 NTC1_MOUSE	Q01705 mus musculus
5	74.5	12.8	2444	1 NTC1_HUMAN	P46531 homo sapien
6	74	12.7	124	1 CHH2_BONMO	P05687 bombyx mori
7	74	12.7	2531	1 NTC1_RAT	Q07008 rattus norv
8	73.5	12.7	660	1 PRIA_BORBU	Q45032 borrelia bu
9	73	12.6	1049	1 ADP1_YEAST	P25371 saccharomyc
10	73	12.6	1700	1 BAR3_CHITE	Q03376 chironomus
11	73	12.6	2437	1 NOTC_BRARE	P46530 brachydanio
12	72	12.4	1193	1 LMG2_HUMAN	Q13753 homo sapien
13	72	12.4	2318	1 NTC3_MOUSE	Q61982 mus musculus
14	71.5	12.3	2703	1 NOTC_DROME	P07207 drosophila
15	71.5	12.3	519	1 TYR2_HUMAN	P40126 homo sapien
16	71.5	12.3	2476	1 ZAN_PIG	Q28983 sus scrofa
17	71	12.2	2907	1 FBN2_MOUSE	Q61555 mus musculus
18	71	12.2	72	1 MT13_MYED	P80248 mytilus edu
19	71	12.2	397	1 PRLA_LYSEN	P00778 lysobacter
20	71	12.2	1134	1 TIE1_MOUSE	Q06806 mus musculus
21	70	12.0	2871	1 FBN1_MOUSE	Q61584 mus musculus
22	69.5	12.0	2524	1 NOTC_XENLA	P21783 xenopus lae
23	69	11.9	2715	1 G156_PAPPR	P13837 paramecium
24	68.5	11.8	581	1 IRR_RAT	Q64716 rattus norv
25	68	11.7	72	1 MT11_MYED	P80246 mytilus edu
26	68	11.7	458	1 PRTC_RABIT	Q28661 oryctolagus
27	68	11.7	1138	1 TIE1_HUMAN	P35590 homo sapien
28	67	11.5	372	1 AGL1_ORTDI	P11218 rattus norv
29	67	11.5	1801	1 LMB2_RAT	P15800 rattus norv
30	67	11.5	72	1 MT12_MYED	P80247 mytilus edu
31	67	11.5	364	1 PGLR_COCCA	P26215 cochlidiobol
32	67	11.5	1136	1 TIE1_BOVIN	Q06805 bos taurus
33	66.5	11.4	2911	1 FBN2_HUMAN	P35556 homo sapien
34	66.5	11.4	182	1 HPRT_RHOCA	P37171 rhodobacter
35	66.5	11.4	71	1 MT21_MYED	P80251 mytilus edu
36	66.5	11.4	444	1 SHU4_ECOLI	P09748 escherichia
37	66.5	11.4	1416	1 YN81_CAEEL	Q03610 caenorhabdi
38	66	11.4	177	1 BTC_MOUSE	Q05928 mus musculus
39	66	11.4	2871	1 FBN1_HUMAN	P35555 homo sapien
40	66	11.4	1300	1 IRR_CAVPO	P14617 cavia porce
41	66	11.4	463	1 MFGM_MOUSE	P21956 mus musculus
42	66	11.4	72	1 MT14_MYED	P80249 mytilus edu
43	66	11.4	309	1 PNAD_PIG	Q28955 sus scrofa

## ALIGNMENTS

RESULT 1  
CO6\_HUMAN  
ID CO6\_HUMAN STANDARD; PRT; 934 AA.  
AC P13671.  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE COMPLEMENT COMPONENT C6 PRECURSOR.  
GN C6.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-31 AND 633-640.  
RX MEDLINE; 90036879.  
RA HAEFLIGER J. A., TSCHOPP J., VIAL N., JENNE D. E.;  
RT "Complete primary structure and functional characterization of the  
sixth component of the human complement system. Identification of the  
C5b-binding domain in complement C6.";  
RL J. BIOL. CHEM. 264:18041-18051(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89380223.  
RA DISCIPPIO R. G., HUGLI T. E.;  
RT "The molecular architecture of human complement component C6.";  
RL J. BIOL. CHEM. 264:16197-16206(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE-BLOOD;  
RA HOBART M. J., FERNIE B., DISCIPPIO R. G.;  
RT "Structure of the human C6 gene.";  
RL BIOCHEMISTRY 32:6198-6205(1993).  
RN [4]  
RP SEQUENCE OF 1-491 FROM N.A.  
RX MEDLINE; 89202413.  
RA CHAKRAVARTI D. N., CHAKRAVARTI B., PARRA C. A., MUELLER-EBERHARD H. J.;  
RT "Structural homology of complement protein C6 with other  
channel-forming proteins of complement.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:2799-2803(1989).  
RN [5]  
RP VARIANT ALLOTYPES C6 A AND B.  
RX MEDLINE; 93326158.  
RA DEWALD G., NOTHEN M. M., CICHON S.;  
RT "Polymorphism of human complement component C6: an amino acid  
substitution (Glu/Ala) within the second thrombospondin repeat  
differentiates between the two common allotypes C6 A and C6 B.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 194:458-464(1993).  
CC -1- FUNCTION: INVOLVED IN THE FORMATION OF THE LYTIC C5B-9M COMPLEX.  
CC -1- DOMAIN: C6 CONTAINS NINE DISTINCT STRUCTURAL UNITS (CYSTEINE-RICH)  
CC -1- WHICH HAVE THE CHARACTERISTIC FEATURES OF MODULES.  
CC -1- PFM: ALL CYSTEINE RESIDUES ARE ASSUMED TO BE CROSS-LINKED TO ONE  
ANOTHER. INDIVIDUAL MODULES CONTAINING AN EVEN NUMBER OF CONSERVED  
CYSTEINE RESIDUES ARE SUPPOSED TO HAVE DISULFIDE LINKAGES ONLY  
WITHIN THE SAME MODULE.  
CC -1- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPIC C6 A.  
CC -1- SIMILARITY: TO COMPLEMENT FACTORS C7, C8, C9, AND TO PERFORIN.  
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way

84

none  
have

44 65.5 11.3 71 1 MT22\_MYED P80252 mytilus edu  
45 65.5 11.3 66 1 MT\_ARIAR P55946 arianta arb

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; J05064; G179704; -
DR DR EMBL; J05024; G307228; -
DR DR EMBL; X72177; G825633; -
DR DR EMBL; J04506; G618466; -
DR DR PIR; A32109; A32109.
DR DR PIR; A34235; A34235.
DR DR PIR; A34372; A34372.
DR DR MIN; 217050; -
DR DR PROSITE; PS00279; MAC_PERFORIN; 1.
DR DR PROSITE; PS00022; EGF_1; 1.
DR DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR DR PROSITE; PS01209; LDLRA_1; 1.
DR DR PROSITE; PS50068; LDLRA_2; 1.
DR DR PFAM; PF00050; kazal; 1.
DR DR PFAM; PF00057; ldl_recept_a; 1.
DR DR PFAM; PF00084; sushi; 2.
DR DR PFAM; PF00090; tsp_1; 3.
DR DR HSP; P01130; 1A1J.
KW KW COMPLEMENT PATHWAY; GLYCOPROTEIN; PLASMA; MEMBRANE ATTACK COMPLEX;
KW CYTOLYSIS; SUSHI; REPEAT; SIGNAL; POLYMORPHISM; EGF-LIKE DOMAIN;
KW TRANSMEMBRANE.
FT FT SIGNAL 1 21
FT FT CHAIN 22 934 COMPLEMENT COMPONENT C6.
FT FT REPEAT 22 80 TYPE-1 TSP 1.
FT FT REPEAT 81 137 TYPE-1 TSP 2.
FT FT DOMAIN 138 175 LDL-RECEPTOR CLASS A.
FT FT TRANSMEM 331 349 POTENTIAL.
FT FT TRANSMEM 354 373 POTENTIAL.
FT FT DOMAIN 517 553 EGF-LIKE.
FT FT REPEAT 562 611 EGF-LIKE.
FT FT DOMAIN 642 934 C5B-BINDING DOMAIN.
FT FT REPEAT 643 700 2 X SUSHI (SCR) REPEATS.
FT FT REPEAT 703 762 SUSHI 1.
FT FT DOMAIN 766 840 SUSHI 2.
FT FT DOMAIN 858 934 COMPLEMENT CONTROL FACTOR I MODULE.
FT FT DOMAIN 140 151 COMPLEMENT CONTROL FACTOR I MODULE.
FT FT DISULFID 146 164 BY SIMILARITY.
FT FT DISULFID 158 173 BY SIMILARITY.
FT FT DISULFID 399 420 BY SIMILARITY.
FT FT DISULFID 523 539 BY SIMILARITY.
FT FT DISULFID 526 541 BY SIMILARITY.
FT FT DISULFID 543 552 BY SIMILARITY.
FT FT DISULFID 644 686 BY SIMILARITY.
FT FT DISULFID 672 699 BY SIMILARITY.
FT FT DISULFID 704 746 BY SIMILARITY.
FT FT DISULFID 732 761 BY SIMILARITY.
FT FT CARBOHYD 324 324 POTENTIAL.
FT FT CARBOHYD 855 855 POTENTIAL.
FT FT VARIANT 119 119 E -> A (IN ALLOTYPE C6 B).
SQ SEQUENCE 934 AA; 104843 MW; D03BBD9D CRC32;

Query Match 14.8%; Score 86; DB 1; Length 934;
Best Local Similarity 29.4%; Pred. NO. 0.33;
Matches 30; Conservative 10; Mismatches 36; Indels 26; Gaps 7;

QY 22 CSMEAFINERIVAGSLIFRAISSIGLECSQVTSRGDLATCPRGFAVTG---CTCGSAC 78
Db 699 CORTCIKPVQVEVLTITPFQRLYRIG-----LTCPKGFVAGPSRYTCQG-- 749
QY 79 GSW--DVRAETTC-----HCQACAGMDWTGARCCRVQP 108
Db 749 NSWTPPIINSLTCEKDXLTKLKGHCQ-LGQKQSGECICMSP 789

RESULT 2
PRTC_BOVIN
ID PRTC_BOVIN STANDARD; PRT; 456 AA.

```

```

AC P00745;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 15-DEC-1988 (REL. 37, LAST ANNOTATION UPDATE)
DE VITAMIN-K DEPENDENT PROTEIN C PRECURSOR (EC 3.4.21.69)
DE (AUTOPROTHROMBIN IIA) (ANTICOAGULANT PROTEIN C) (FRAGMENT).
GN PROC.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85014826.
RA LONG G.L., BALAGAJE R.M., MCGILLIVRAY R.T.A.;
RT "Cloning and sequencing of liver cDNA coding for bovine protein C.";
RL PROC. NATL. ACAD. SCI. U.S.A. 81:5653-5656(1984).
RN [2]
RP SEQUENCE OF 40-194.
RX MEDLINE; 83007325.
RA FERNLUND P., STENFLO J.;
RT "Amino acid sequence of the light chain of bovine protein C.";
RL J. BIOL. CHEM. 257:12170-12179(1982).
RN [3]
RP REVISION TO 110.
RX MEDLINE; 83169769.
RA DRAKENBERG T., FERNLUND P., ROEPSTORFF P., STENFLO J.;
RT "Beta-hydroxyaspartic acid in vitamin K-dependent protein C.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1802-1806(1983).
RN [4]
RP SEQUENCE OF 197-456.
RX MEDLINE; 83007326.
RA STENFLO J., FERNLUND P.;
RT "Amino acid sequence of the heavy chain of bovine protein C.";
RL J. BIOL. CHEM. 257:12180-12190(1982).
RN [5]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE; 83213513.
RA ESMON N.L., DEBAULT L.E., ESMON C.T.;
RT "Proteolytic formation and properties of gamma-carboxyglutamic acid-
domainless protein C.";
RL J. BIOL. CHEM. 258:5548-5553(1983).
RN [6]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE; 83213514.
RA JOHNSON A.E., ESMON N.L., LAUE T.M., ESMON C.T.;
RT "Structural changes required for activation of protein C are induced
by Ca2+ binding to a high affinity site that does not contain gamma-
carboxyglutamic acid.";
RL J. BIOL. CHEM. 258:5554-5560(1983).
CC -!- FUNCTION: PROTEIN C IS A VITAMIN K-DEPENDENT SERINE PROTEASE THAT
REGULATES BLOOD COAGULATION BY INACTIVATING FACTORS VA AND VIIIA
IN THE PRESENCE OF CALCIUM IONS AND PHOSPHOLIPIDS.
CC -!- CATALYTIC ACTIVITY: DEGRADATION OF BLOOD COAGULATION FACTORS VA
AND VIIIA.
CC -!- SUBUNIT: SYNTHESIZED AS A SINGLE CHAIN PRECURSOR, WHICH IS CLEAVED
INTO A LIGHT CHAIN AND A HEAVY CHAIN HELD TOGETHER BY A DISULFIDE
BOND. THE ENZYME IS THEN ACTIVATED BY THROMBIN, WHICH CLEAVES A
TETRAPEPTIDE FROM THE AMINO END OF THE HEAVY CHAIN; THIS
REACTION, WHICH OCCURS AT THE SURFACE OF ENDOTHELIAL CELLS, IS
STRONGLY PROMOTED BY THROMBOMODULIN.
CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -!- PTM: THE VITAMIN K-DEPENDENT, ENZYMIC CARBOXYLATION OF SOME
GLU RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND CALCIUM.
CC -!- CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO ANOTHER SITE, BEYOND
THE GLA DOMAIN. THIS GLA-INDEPENDENT BINDING SITE IS NECESSARY FOR
THE RECOGNITION OF THE THROMBIN-THROMBOMODULIN COMPLEX.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```



FT	DOMAIN	348	EGF-LIKE 3.	FT	DISULFID	849	890	BY SIMILARITY.
FT	DOMAIN	425	EGF-LIKE 4.	FT	DISULFID	892	901	BY SIMILARITY.
FT	DOMAIN	427	EGF-LIKE 5.	FT	DISULFID	908	919	BY SIMILARITY.
FT	DOMAIN	464	EGF-LIKE 6.	FT	DISULFID	913	928	BY SIMILARITY.
FT	DOMAIN	501	EGF-LIKE 7.	FT	DISULFID	930	939	BY SIMILARITY.
FT	DOMAIN	532	EGF-LIKE 8.	FT	DISULFID	946	957	BY SIMILARITY.
FT	DOMAIN	545	EGF-LIKE 9.	FT	DISULFID	952	966	BY SIMILARITY.
FT	DOMAIN	582	EGF-LIKE 10.	FT	DISULFID	968	977	BY SIMILARITY.
FT	DOMAIN	609	EGF-LIKE 11.	FT	DISULFID	984	995	BY SIMILARITY.
FT	DOMAIN	648	EGF-LIKE 12.	FT	DISULFID	989	1009	BY SIMILARITY.
FT	DOMAIN	687	EGF-LIKE 13.	FT	DISULFID	1011	1020	BY SIMILARITY.
FT	DOMAIN	725	EGF-LIKE 14.	FT	DISULFID	1211	1222	BY SIMILARITY.
FT	DOMAIN	763	EGF-LIKE 15.	FT	DISULFID	1216	1231	BY SIMILARITY.
FT	DOMAIN	802	EGF-LIKE 16.	FT	DISULFID	1233	1242	BY SIMILARITY.
FT	DOMAIN	840	EGF-LIKE 17.	FT	DISULFID	1485	1496	BY SIMILARITY.
FT	DOMAIN	904	EGF-LIKE 18.	FT	DISULFID	1490	1505	BY SIMILARITY.
FT	DOMAIN	942	EGF-LIKE 19.	FT	DISULFID	1507	1516	BY SIMILARITY.
FT	DOMAIN	980	EGF-LIKE 20.	FT	DISULFID	1763	1774	BY SIMILARITY.
FT	DOMAIN	1207	EGF-LIKE 21.	FT	DISULFID	1768	1783	BY SIMILARITY.
FT	DOMAIN	1481	EGF-LIKE 22.	FT	DISULFID	1785	1794	BY SIMILARITY.
FT	DOMAIN	1759	EGF-LIKE 23.	FT	DISULFID	1801	1812	BY SIMILARITY.
FT	DOMAIN	1833	EGF-LIKE 24.	FT	DISULFID	1806	1821	BY SIMILARITY.
FT	DOMAIN	1835	EGF-LIKE 25.	FT	DISULFID	1833	1852	BY SIMILARITY.
FT	DOMAIN	1874	EGF-LIKE 26.	FT	DISULFID	1839	1850	BY SIMILARITY.
FT	DOMAIN	1915	EGF-LIKE 27.	FT	DISULFID	1844	1859	BY SIMILARITY.
FT	DOMAIN	1933	EGF-LIKE 28.	FT	DISULFID	1861	1870	BY SIMILARITY.
FT	DOMAIN	1951	EGF-LIKE 29.	FT	DISULFID	1878	1889	BY SIMILARITY.
FT	DOMAIN	1991	BY SIMILARITY.	FT	DISULFID	1883	1903	BY SIMILARITY.
FT	DOMAIN	2030	BY SIMILARITY.	FT	DISULFID	1905	1914	BY SIMILARITY.
FT	DOMAIN	271	BY SIMILARITY.	FT	DISULFID	1919	1930	BY SIMILARITY.
FT	DOMAIN	276	BY SIMILARITY.	FT	DISULFID	1924	1939	BY SIMILARITY.
FT	DOMAIN	291	BY SIMILARITY.	FT	DISULFID	1941	1950	BY SIMILARITY.
FT	DOMAIN	293	BY SIMILARITY.	FT	DISULFID	1957	1968	BY SIMILARITY.
FT	DOMAIN	310	BY SIMILARITY.	FT	DISULFID	1962	1977	BY SIMILARITY.
FT	DOMAIN	321	BY SIMILARITY.	FT	DISULFID	1979	1988	BY SIMILARITY.
FT	DOMAIN	331	BY SIMILARITY.	FT	DISULFID	1995	2008	BY SIMILARITY.
FT	DOMAIN	333	BY SIMILARITY.	FT	DISULFID	2002	2017	BY SIMILARITY.
FT	DOMAIN	335	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	352	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	363	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	374	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	376	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	385	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	392	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	397	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	412	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	424	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	431	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	436	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	453	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	462	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	468	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	473	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	490	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	505	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	509	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	522	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	549	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	556	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	571	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	580	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	586	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	591	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	604	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	613	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	618	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	634	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	645	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	664	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	684	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	691	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	702	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	711	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	722	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	740	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	749	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	751	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	760	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	777	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	787	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	799	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	806	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	811	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	826	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	837	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.
FT	DOMAIN	844	BY SIMILARITY.	FT	DISULFID	2019	2028	BY SIMILARITY.

Query Match 13.0%; Score 75.5; DB 1; Length 2139;  
Best Local Similarity 25.0%; Pred. No. 7.5;  
Matches 24; Conservative 10; Mismatches 37; Indels 25; Gaps 4;  
QY 16 VSSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTSRGLATCPRGFAVTGCTG 75  
DB 346 VNIHPLCQTNPCNLNGACVVG-----GSGALTCE-----CPKGYAGARCEVD 388  
QY 76 S-ACGSMDVRAETTC-----HCQCAGMDWTGARC 103



DR .PROSITE: PS01186; EGF\_2; 26.  
DR PROSITE: PS01187; EGF\_CA; 18.  
DR PFAM: PF00008; EGF; 35.  
DR PFAM: PF00023; ank; 6.  
DR PFAM: PF00066; notch; 3.  
DR HSP: P00740; IIXA.  
KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;  
KW TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.  
FT SIGNAL 1 18  
FT CHAIN 19 >2444 POTENTIAL.  
FT DOMAIN 19 1736 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.  
FT DOMAIN 19 1736 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1737 1757 POTENTIAL.  
FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 20 58 EGF-LIKE 1.  
FT DOMAIN 59 99 EGF-LIKE 2.  
FT DOMAIN 102 139 EGF-LIKE 3.  
FT DOMAIN 140 176 EGF-LIKE 4.  
FT DOMAIN 178 216 EGF-LIKE 5.  
FT DOMAIN 218 255 EGF-LIKE 6.  
FT DOMAIN 257 293 EGF-LIKE 7.  
FT DOMAIN 295 333 EGF-LIKE 8.  
FT DOMAIN 335 371 EGF-LIKE 9.  
FT DOMAIN 372 410 EGF-LIKE 10.  
FT DOMAIN 412 450 EGF-LIKE 11.  
FT DOMAIN 452 488 EGF-LIKE 12.  
FT DOMAIN 490 526 EGF-LIKE 13.  
FT DOMAIN 528 564 EGF-LIKE 14.  
FT DOMAIN 566 601 EGF-LIKE 15.  
FT DOMAIN 603 639 EGF-LIKE 16.  
FT DOMAIN 641 676 EGF-LIKE 17.  
FT DOMAIN 678 714 EGF-LIKE 18.  
FT DOMAIN 716 751 EGF-LIKE 19.  
FT DOMAIN 753 789 EGF-LIKE 20.  
FT DOMAIN 791 827 EGF-LIKE 21.  
FT DOMAIN 829 868 EGF-LIKE 22.  
FT DOMAIN 870 906 EGF-LIKE 23.  
FT DOMAIN 908 944 EGF-LIKE 24.  
FT DOMAIN 946 982 EGF-LIKE 25.  
FT DOMAIN 984 1020 EGF-LIKE 26.  
FT DOMAIN 1022 1058 EGF-LIKE 27.  
FT DOMAIN 1060 1096 EGF-LIKE 28.  
FT DOMAIN 1098 1144 EGF-LIKE 29.  
FT DOMAIN 1146 1182 EGF-LIKE 30.  
FT DOMAIN 1184 1220 EGF-LIKE 31.  
FT DOMAIN 1222 1266 EGF-LIKE 32.  
FT DOMAIN 1268 1306 EGF-LIKE 33.  
FT DOMAIN 1308 1347 EGF-LIKE 34.  
FT DOMAIN 1349 1385 EGF-LIKE 35.  
FT DOMAIN 1388 1427 EGF-LIKE 36.  
FT DOMAIN 1446 1563 3 X LIN/NOTCH REPEATS.  
FT REPEAT 1446 1481 LIN/NOTCH 1.  
FT REPEAT 1482 1523 LIN/NOTCH 2.  
FT REPEAT 1524 1563 LIN/NOTCH 3.  
FT DOMAIN 1876 2087 6 X ANK MOTIF REPEATS.  
FT REPEAT 1876 1921 ANK MOTIF 1.  
FT REPEAT 1923 1954 ANK MOTIF 2.  
FT REPEAT 1956 1987 ANK MOTIF 3.  
FT REPEAT 1990 2021 ANK MOTIF 4.  
FT REPEAT 2023 2054 ANK MOTIF 5.  
FT REPEAT 2056 2087 ANK MOTIF 6.  
FT DOMAIN 1576 1579 POLY-VAL.  
FT DOMAIN 1662 1665 POLY-ARG.  
FT DOMAIN 1729 1732 POLY-PRO.  
FT DOMAIN 1741 1744 POLY-ALA.  
FT DOMAIN 1902 1905 POLY-GLU.  
FT DOMAIN 2260 2263 POLY-GLY.  
FT DOMAIN 2404 2407 POLY-GLN.  
FT DOMAIN 2411 2418 POLY-PRO.  
FT DISULFID 24 37 BY SIMILARITY.  
FT DISULFID 31 46 BY SIMILARITY.  
FT DISULFID 48 57 BY SIMILARITY.  
FT DISULFID 63 74 BY SIMILARITY.  
FT DISULFID 68 87 BY SIMILARITY.  
FT DISULFID 89 106 FT DISULFID  
FT DISULFID 106 117 FT DISULFID  
FT DISULFID 111 127 FT DISULFID  
FT DISULFID 129 138 FT DISULFID  
FT DISULFID 144 155 FT DISULFID  
FT DISULFID 149 164 FT DISULFID  
FT DISULFID 166 175 FT DISULFID  
FT DISULFID 182 195 FT DISULFID  
FT DISULFID 189 204 FT DISULFID  
FT DISULFID 206 215 FT DISULFID  
FT DISULFID 222 233 FT DISULFID  
FT DISULFID 227 243 FT DISULFID  
FT DISULFID 245 254 FT DISULFID  
FT DISULFID 261 272 FT DISULFID  
FT DISULFID 266 281 FT DISULFID  
FT DISULFID 283 292 FT DISULFID  
FT DISULFID 289 312 FT DISULFID  
FT DISULFID 306 321 FT DISULFID  
FT DISULFID 323 332 FT DISULFID  
FT DISULFID 339 350 FT DISULFID  
FT DISULFID 344 359 FT DISULFID  
FT DISULFID 361 370 FT DISULFID  
FT DISULFID 376 387 FT DISULFID  
FT DISULFID 381 398 FT DISULFID  
FT DISULFID 400 409 FT DISULFID  
FT DISULFID 416 429 FT DISULFID  
FT DISULFID 423 438 FT DISULFID  
FT DISULFID 440 449 FT DISULFID  
FT DISULFID 456 467 FT DISULFID  
FT DISULFID 461 476 FT DISULFID  
FT DISULFID 478 487 FT DISULFID  
FT DISULFID 494 505 FT DISULFID  
FT DISULFID 499 514 FT DISULFID  
FT DISULFID 516 525 FT DISULFID  
FT DISULFID 532 543 FT DISULFID  
FT DISULFID 537 552 FT DISULFID  
FT DISULFID 554 563 FT DISULFID  
FT DISULFID 570 580 FT DISULFID  
FT DISULFID 575 589 FT DISULFID  
FT DISULFID 591 600 FT DISULFID  
FT DISULFID 607 618 FT DISULFID  
FT DISULFID 612 627 FT DISULFID  
FT DISULFID 629 638 FT DISULFID  
FT DISULFID 645 655 FT DISULFID  
FT DISULFID 660 664 FT DISULFID  
FT DISULFID 666 675 FT DISULFID  
FT DISULFID 682 693 FT DISULFID  
FT DISULFID 687 702 FT DISULFID  
FT DISULFID 704 713 FT DISULFID  
FT DISULFID 720 730 FT DISULFID  
FT DISULFID 725 739 FT DISULFID  
FT DISULFID 741 750 FT DISULFID  
FT DISULFID 757 768 FT DISULFID  
FT DISULFID 762 777 FT DISULFID  
FT DISULFID 779 788 FT DISULFID  
FT DISULFID 795 806 FT DISULFID  
FT DISULFID 800 815 FT DISULFID  
FT DISULFID 817 826 FT DISULFID  
FT DISULFID 833 844 FT DISULFID  
FT DISULFID 838 855 FT DISULFID  
FT DISULFID 857 867 FT DISULFID  
FT DISULFID 874 885 FT DISULFID  
FT DISULFID 879 894 FT DISULFID  
FT DISULFID 896 905 FT DISULFID  
FT DISULFID 912 923 FT DISULFID  
FT DISULFID 917 932 FT DISULFID  
FT DISULFID 934 943 FT DISULFID  
FT DISULFID 948 959 FT DISULFID  
FT DISULFID 993 1008 FT DISULFID  
FT DISULFID 1010 1019 FT DISULFID  
FT DISULFID 1026 1037 FT DISULFID  
FT DISULFID 1031 1046 FT DISULFID  
FT DISULFID 1048 1057 FT DISULFID





```
FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1097 1143 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1307 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1348 1384 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1387 1426 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1449 1462 CYS-RICH.
FT DOMAIN 1865 2076 6 X ANK MOTIF REPEATS.
FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2076 ANK MOTIF 6.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
```

```
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.
FT DISULFID 1133 1142 BY SIMILARITY.
FT DISULFID 1149 1160 BY SIMILARITY.
FT DISULFID 1154 1169 BY SIMILARITY.
FT DISULFID 1171 1180 BY SIMILARITY.
FT DISULFID 1187 1198 BY SIMILARITY.
FT DISULFID 1192 1207 BY SIMILARITY.
FT DISULFID 1209 1218 BY SIMILARITY.
FT DISULFID 1225 1244 BY SIMILARITY.
FT DISULFID 1238 1253 BY SIMILARITY.
FT DISULFID 1255 1264 BY SIMILARITY.
FT DISULFID 1271 1284 BY SIMILARITY.
FT DISULFID 1276 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1311 1322 BY SIMILARITY.
FT DISULFID 1316 1334 BY SIMILARITY.
FT DISULFID 1336 1345 BY SIMILARITY.
FT DISULFID 1352 1363 BY SIMILARITY.
FT DISULFID 1357 1372 BY SIMILARITY.
FT DISULFID 1374 1383 BY SIMILARITY.
FT DISULFID 1391 1403 BY SIMILARITY.
```

Query Match 12.7%; Score 74; DB 1; Length 2531;

Best Local Similarity 25.7%; Pred. No. 12;

Matches 36; Conservative 12; Mismatches 52; Indels 40; Gaps 8;

QY 4 LCLLLPVL---GLLVSKTLCSMEEAINERIQVAGSLIFRAISS---IGLECCS----- 54

Db 9 LCLTLLPALAARGLSQPS-----GTCINGRCRCEVANGTEACVSGAFVGORCDPSPCL 64

QY 54 -----VTSRGDL----ATCPRGFVATGC--TCGSAC-----GSDVRAETT 88

Db 65 STPCNAGTCVVDHGGIVDYACSPGLGFSPLCLTPLANACLANPCRNCGTCDLLTITE 124  
QY 89 CHCQAGMDWTGACRCRVQP 108  
Db 125 YKCRCP-PGWSGKSCQADP 143

RESULT 8  
PRIA\_BORBU  
ID PRIA\_BORBU STANDARD; PRT; 660 AA.  
AC Q45032; 051047;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).  
GN PRIA OR BB0014.  
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).  
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HB19;  
RA BOURSAUX-EUDE C.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / B31;  
RX MEDLINE; 98065943.  
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
RA DOUGHERTY B., TOMB J.F., FLEISCHMANN R.D., RICHARDSON D.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENER J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";  
RL NATURE 390:580-586(1997).  
CC -1- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X97449; E238797; -  
DR EMBL; AE001115; G2687882; -  
DR TIGR; BB0014; -  
DR PFAM; PF00271; helicase\_C; 1.  
KW DNA REPLICATION; DNA-BINDING; ATP-BINDING; HELICASE; PRIMOSOME;  
KZ ZINC-FINGER.  
FT NP\_BIND 158 165 ATP (POTENTIAL).  
FT SITE 256 259 DEEH BOX.  
FT 2N\_FING 370 382 C4-TYPE (POTENTIAL).  
FT 2N\_FING 397 413 C4-TYPE (POTENTIAL).  
FT CONFLICT 126 126 P -> L (IN REF. 1).  
FT CONFLICT 555 555 D -> N (IN REF. 1).  
SQ SEQUENCE 660 AA; 77551 MW; B799DC6E CRC32;

Query Match 12.7%; Score 73.5; DB 1; Length 660;  
Best Local Similarity 29.2%; Pred. No. 4.1;  
Matches 28; Conservative 13; Mismatches 26; Indels 29; Gaps 7;

QY 16 VSSKTLCSMEERINRIQEVAGSLIFRAISSIG-----LECSQVTSRGDLATCPRGFAVIG 71  
Db 337 ISSELSYIQKSLNKRQ-----SLIF--INKRGLKLECNEC---GHICCP-----N 381

QY 72 CTCG-----SAGSGWDVRAETTCCHC-QCAGMD 97  
Db 382 CSFGLIYHKKNKLLCHYCSYKTKTASHCPQCESKD 417

RESULT 9  
ADP1\_YEAST  
ID ADP1\_YEAST STANDARD; PRT; 1049 AA.  
AC P25371;  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR.  
GN ADP1 OR YCR011C OR YCR11C OR YCR105.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACHAROMYCETALES;  
OC SACHAROMYCETACEAE; SACHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92160395.  
RA PURNELLE B., SKALA J., GOFFEAU A.;  
RT "The product of the YCR105 gene located on the chromosome III from Saccharomyces cerevisiae presents homologies to ATP-dependent permeases";  
RL YEAST 7:867-872(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92327849.  
RA SKALA J., PURNELLE B., GOFFEAU A.;  
RT "The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading frames including the RVS161, ADP1 and PGK genes.";  
RL YEAST 8:409-417(1992).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X59720; E264475; -  
DR PIR; S19421; S19421.  
DR PIR; S40914; S40914.  
DR SGD; L0000049; ADP1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_Tran; 1.  
DR PFAM; PF00008; EGF; 1.  
KW ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN; TRANSPORT; SIGNAL.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.  
FT NP\_BIND 423 430 ATP (BY SIMILARITY).  
FT TRANSMEM 325 345 POTENTIAL.  
FT TRANSMEM 464 481 POTENTIAL.  
FT TRANSMEM 794 814 POTENTIAL.  
FT TRANSMEM 829 849 POTENTIAL.  
FT TRANSMEM 878 898 POTENTIAL.  
FT TRANSMEM 910 930 POTENTIAL.  
FT TRANSMEM 938 958 POTENTIAL.  
FT TRANSMEM 1001 1021 POTENTIAL.  
FT TRANSMEM 1025 1045 POTENTIAL.  
FT CARBOHYD 50 50 POTENTIAL.  
FT CARBOHYD 114 114 POTENTIAL.  
FT CARBOHYD 165 165 POTENTIAL.  
FT CARBOHYD 221 221 POTENTIAL.





## RESULT 12

LMG2\_HUMAN  
ID LMG2\_HUMAN STANDARD: PRT: 1193 AA.  
AC Q13753; Q13752; Q14941; Q02536; Q02537;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DI LAMININ GAMMA-2 CHAIN PRECURSOR.  
GN LAMC2.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RC TISSUE-PLACENTA;  
RX MEDLINE: 96230326.  
RA AIRENNE T., HAKKANA H., SAINIO K., KALLUNKI T., KALLUNKI P.,  
RA SARIOLO H., TRYGGVASON K.;  
RT "Structure of the human laminin gamma 2 chain gene (LAMC2):  
RT alternative splicing with different tissue distribution of two  
RT transcripts.";  
RL GENOMICS 32:54-64(1996).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1090-1114.  
RC TISSUE-EPIDERMIS, AND KERATINOCYTES;  
RX MEDLINE: 94139694.  
RA VAILLY J., VERRANDO P., CHAMPLAUD M.F., GERECKE D., WAGMAN D.W.,  
RA BAUDOUIN C., ABERDAM D., BURGESSON R., BAUER E., ORTONE J.P.;  
RT "The 100-kDa chain of nidogen/kallinin is a laminin B2 chain variant.";  
RL EUR. J. BIOCHEM. 219:209-218(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-FIBROSARCOMA;  
RX MEDLINE: 93016279.  
RA KALLUNKI P., SAINIO K., EDDY R., BYERS M., KALLUNKI T.,  
RA SARIOLO H., BECK K., HIRVONEN H., SHOWS T.B., TRYGGVASON K.;  
RT "A truncated laminin homologous to the B2 chain: structure,  
RT spatial expression, and chromosomal assignment.";  
RL J. CELL BIOL. 119:679-693(1992).  
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, AND ORGANIZATION OF  
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
CC -!- SUBUNIT: THE GAMMA-2 CHAIN IS A SUBUNIT OF LAMININ-5 (EPILIGRIN/  
CC KALININ/NICEIN).  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT  
CC MEMBRANES (MAJOR COMPONENT).  
CC -!- TISSUE SPECIFICITY: THE LARGE VARIANT IS EXPRESSED ONLY IN  
CC SPECIFIC EPITHELIAL CELLS OF EMBRYONIC AND NEONATAL TISSUES. IN  
CC 17-WEEK OLD EMBRYO THE SMALL VARIANT IS FOUND IN CEREBRAL CORTEX,  
CC LUNG, AND DISTAL TUBES OF KIDNEY, BUT NOT IN EPITHELIA EXCEPT FOR  
CC DISTAL TUBULE.  
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -!- DOMAIN: DOMAIN IV IS GLOBULAR.  
CC -!- DISEASE: JUNCTIONAL EPIDERMOLYSIS BULLOSA GRAVIS IS A BLISTERING  
CC DISORDER IN SKIN THAT IS CHARACTERIZED BY A SEPARATION OF BASAL  
CC CELLS FROM THE BASEMENT MEMBRANE DUE TO A DECREASED NUMBER OF  
CC HEMIDESMOSOMES. LAMININ-5 IS MISSING FROM THE BASEMENT MEMBRANE OF  
CC PATIENTS WITH THE GRAVIS FORM OF EPIDERMOLYSIS BULLOSA.  
CC -!- ALTERNATIVE PRODUCTS: THERE EXIST A SMALL AND A LARGE VARIANT,  
CC WHICH DIFFER IN THEIR C-TERMINAL, DUE TO ALTERNATIVE SPLICING OF  
CC THE SAME GENE.  
CC -!- SIMILARITY: CONTAINS 7.5 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U31201; G1236323; JOINED.  
CC EMBL; U31178; G1236323; JOINED.  
CC EMBL; U31179; G1236323; JOINED.  
CC EMBL; U31180; G1236323; JOINED.  
CC EMBL; U31181; G1236323; JOINED.  
CC EMBL; U31182; G1236323; JOINED.  
CC EMBL; U31183; G1236323; JOINED.  
CC EMBL; U31184; G1236323; JOINED.  
CC EMBL; U31186; G1236323; JOINED.  
CC EMBL; U31187; G1236323; JOINED.  
CC EMBL; U31188; G1236323; JOINED.  
CC EMBL; U31189; G1236323; JOINED.  
CC EMBL; U31190; G1236323; JOINED.  
CC EMBL; U31191; G1236323; JOINED.  
CC EMBL; U31192; G1236323; JOINED.  
CC EMBL; U31193; G1236323; JOINED.  
CC EMBL; U31194; G1236323; JOINED.  
CC EMBL; U31195; G1236323; JOINED.  
CC EMBL; U31196; G1236323; JOINED.  
CC EMBL; U31197; G1236323; JOINED.  
CC EMBL; U31198; G1236323; JOINED.  
CC EMBL; U31199; G1236323; JOINED.  
CC EMBL; U31200; G1236323; JOINED.  
CC EMBL; U31178; G1280520; JOINED.  
CC EMBL; U31179; G1280520; JOINED.  
CC EMBL; U31180; G1280520; JOINED.  
CC EMBL; U31181; G1280520; JOINED.  
CC EMBL; U31182; G1280520; JOINED.  
CC EMBL; U31183; G1280520; JOINED.  
CC EMBL; U31184; G1280520; JOINED.  
CC EMBL; U31186; G1280520; JOINED.  
CC EMBL; U31187; G1280520; JOINED.  
CC EMBL; U31188; G1280520; JOINED.  
CC EMBL; U31189; G1280520; JOINED.  
CC EMBL; U31190; G1280520; JOINED.  
CC EMBL; U31191; G1280520; JOINED.  
CC EMBL; U31192; G1280520; JOINED.  
CC EMBL; U31193; G1280520; JOINED.  
CC EMBL; U31194; G1280520; JOINED.  
CC EMBL; U31195; G1280520; JOINED.  
CC EMBL; U31196; G1280520; JOINED.  
CC EMBL; U31197; G1280520; JOINED.  
CC EMBL; U31198; G1280520; JOINED.  
CC EMBL; X73902; G452755; JOINED.  
CC EMBL; U31008; G34230; JOINED.  
CC EMBL; U31009; G34232; JOINED.  
CC MIM: 150292; EGF\_1; 4.  
CC PROSITE: PS00022; EGF\_2; 2.  
CC PROSITE: PS01186; EGF\_2; 2.  
CC PROSITE: PS01248; LAMININ\_TYPE\_EGF; 6.  
CC PFAM: PF00052; laminin\_B; 1.  
CC PFAM: PF00053; laminin\_EGF; 6.  
CC HSP: P02468; ITLE.  
CC KW GLYCOPROTEIN; BASEMENT MEMBRANE; EXTRACELLULAR MATRIX; COILED COIL;  
CC KW LAMININ EGF-LIKE DOMAIN; CELL ADHESION; REPEAT; SIGNAL;  
CC FT ALTERNATIVE SPLICING.  
CC FT SIGNAL 1 21  
CC FT CHAIN 22 1193  
CC FT DOMAIN 28 196  
CC  
CC FT DOMAIN 28 83  
CC FT DOMAIN 84 130  
CC FT DOMAIN 139 186  
CC FT DOMAIN 187 196  
CC FT DOMAIN 197 381  
CC FT DOMAIN 382 602  
CC FT DOMAIN 382 415  
CC FT DOMAIN 416 461  
CC  
CC POTENTIAL.  
CC LAMININ GAMMA-2 CHAIN.  
CC 3.5 X LAMININ EGF-LIKE REPEATS (DOMAIN  
CC V).  
CC LAMININ EGF-LIKE 1.  
CC LAMININ EGF-LIKE 2.  
CC LAMININ EGF-LIKE 3.  
CC LAMININ EGF-LIKE 4 (N-TERMINAL).  
CC LAMININ DOMAIN IV.  
CC 4 X LAMININ EGF-LIKE REPEATS (DOMAIN V).  
CC LAMININ EGF-LIKE 4 (C-TERMINAL).  
CC LAMININ EGF-LIKE 5.







RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86079539.  
 RA WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;  
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene  
 product that shares homology with proteins containing EGF-like  
 repeats.";  
 RL CELL 43:567-581(1985).  
 RN [2].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE: 87064624.  
 RA KIDD S., KELLEY M.R., YOUNG M.W.;  
 RT "Sequence of the notch locus of *Drosophila melanogaster*: relationship  
 of the encoded protein to mammalian clotting and growth factors.";  
 RL MOL. CELL. BIOL. 6:3094-3108(1986).  
 RN [3].  
 RP SEQUENCE OF 2505-2611 FROM N.A.  
 RX MEDLINE: 85099329.  
 RA WHARTON K.A., YEDVOENICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;  
 RT "opa: a novel family of transcribed repeats shared by the Notch locus  
 and other developmentally regulated loci in *D. melanogaster*.";  
 RL CELL 40:55-62(1985).  
 RN [4].  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RX MEDLINE: 87257846.  
 RA KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;  
 RT "Restriction of P-element insertions at the Notch locus of *Drosophila*  
 melanogaster.";  
 RL MOL. CELL. BIOL. 7:1545-1548(1987).  
 RN [5].  
 RP REVIEW.  
 RA HARRIS W.A.;  
 RT "Many cell types specified by Notch function.";  
 RL CURR. BIOL. 1:120-122(1991).  
 CC CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF  
 ECTODERM.  
 CC CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC CC -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART  
 OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,  
 THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.  
 CC CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC -----  
 DR EMBL: M16152; G157988;  
 DR EMBL: M16153; G157988; JOINED.  
 DR EMBL: M16149; G157988; JOINED.  
 DR EMBL: M16150; G157988; JOINED.  
 DR EMBL: M16151; G157988; JOINED.  
 DR EMBL: K03508; G157993;  
 DR EMBL: M13689; G157993; JOINED.  
 DR EMBL: K03507; G157993; JOINED.  
 DR EMBL: M12175; G950317;  
 DR EMBL: M16025; G157995;  
 DR PIR: A24420; A24420.  
 DR PIR: A24768; A24768.  
 DR PIR: A05267; A05267.  
 DR FLYBASE: FBgn0004647; N.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE: PS00022; EGF\_1; 34.  
 DR PROSITE: PS01186; EGF\_2; 28.  
 DR PROSITE: PS01187; EGF\_CA; 22.  
 DR PFAM: PF00008; EGF; 36.  
 DR PFAM: PF00023; ank; 6.

DR PFAM: PF00066; notch; 3.  
 DR HSSP: P00740; LIXA.  
 KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;  
 KW TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.  
 FT SIGNAL 1 44  
 FT CHAIN 45 2703  
 FT DOMAIN 45 1745  
 FT TRANSMEM 1746 1766  
 FT DOMAIN 1767 2703  
 FT DOMAIN 58 1451  
 FT DOMAIN 58 95  
 FT DOMAIN 96 136  
 FT DOMAIN 139 176  
 FT DOMAIN 177 215  
 FT DOMAIN 217 253  
 FT DOMAIN 255 291  
 FT DOMAIN 293 329  
 FT DOMAIN 331 370  
 FT DOMAIN 372 408  
 FT DOMAIN 409 447  
 FT DOMAIN 449 486  
 FT DOMAIN 488 524  
 FT DOMAIN 526 562  
 FT DOMAIN 564 600  
 FT DOMAIN 602 637  
 FT DOMAIN 639 675  
 FT DOMAIN 677 713  
 FT DOMAIN 715 751  
 FT DOMAIN 753 789  
 FT DOMAIN 791 827  
 FT DOMAIN 829 865  
 FT DOMAIN 867 905  
 FT DOMAIN 907 944  
 FT DOMAIN 946 982  
 FT DOMAIN 984 1020  
 FT DOMAIN 1022 1058  
 FT DOMAIN 1060 1096  
 FT DOMAIN 1136 1181  
 FT DOMAIN 1183 1219  
 FT DOMAIN 1221 1257  
 FT DOMAIN 1259 1295  
 FT DOMAIN 1297 1335  
 FT DOMAIN 1337 1373  
 FT DOMAIN 1375 1412  
 FT DOMAIN 1415 1451  
 FT DOMAIN 1475 1593  
 FT REPEAT 1475 1513  
 FT REPEAT 1514 1553  
 FT REPEAT 1554 1593  
 FT REPEAT 1596 2109  
 FT DOMAIN 2538 2586  
 FT DISULFID 62 73  
 FT DISULFID 67 83  
 FT DISULFID 85 94  
 FT DISULFID 100 111  
 FT DISULFID 105 124  
 FT DISULFID 126 135  
 FT DISULFID 143 154  
 FT DISULFID 148 164  
 FT DISULFID 166 175  
 FT DISULFID 181 192  
 FT DISULFID 186 203  
 FT DISULFID 205 214  
 FT DISULFID 221 232  
 FT DISULFID 226 241  
 FT DISULFID 243 252  
 FT DISULFID 259 270  
 FT DISULFID 284 279  
 FT DISULFID 281 290  
 FT DISULFID 297 308  
 FT DISULFID 302 317  
 FT DISULFID 319 328



FT CARBOHYD 178 178 POTENTIAL.  
 FT CARBOHYD 237 237 POTENTIAL.  
 FT CARBOHYD 300 300 POTENTIAL.  
 FT CARBOHYD 342 342 POTENTIAL.  
 FT CARBOHYD 377 377 POTENTIAL.  
 SQ SEQUENCE 519 AA; 59145 MW; 4FEFCDD2 CRC32;

Query Match 12.3%; Score 71.5; DB 1; Length 519;

Best Local Similarity 27.4%; Pred. No. 5.2;

Matches 26; Conservative 7; Mismatches 25; Indels 37; Gaps 4;

Qy 51 CQSVTSRGDLATCPRGFAVTGCTCGSAG---SWDVRAET----- 88

Db 29 CMTVDLVNKECCPLGAEANVCGSQGQCTEVRAETRPWSGPYILRNODDRELWPR 88

Qy 88 -----TCHC--QCAGMD-----WTGARCCRVQP 108

Db 89 KFFHRTCKTGNFAGYNGCDCKFGWTGPNCKPKP 123

Search completed: October 2, 1999, 11:06:11

Job time: 214 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 10:38:38 ; Search time 12.13 Seconds  
(without alignments)  
356.725 Million cell updates/sec

Title: US-09-099-898-2  
Perfect score: 581  
Sequence: 1 MKALCLLLLPVLGLVSSKT.....CHCOCAGMDWTGARCCRVQP 108

Scoring table: BLOSUM62

Searched: 122810-seqs, 40065486 residues

Database : PIR\_50.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	14.8	934	2 A34372	complement C6 prec
2	82	14.1	2321	2 S78549	notch3 protein - h
3	80.5	13.9	456	1 KXBO	protein C (activat
4	77	13.3	295	2 JC5559	lectin-B - Virgini
5	75.5	13.0	2531	2 A46019	gene Notch-1 prote
6	75.5	13.0	387	2 B49175	Notch A protein -
7	75.5	13.0	2139	2 A35672	Crumbs protein - f
8	75	12.9	124	2 A21761	high-cysteine chor
9	74.5	12.8	2555	2 A40043	notch protein homo
10	74	12.7	2531	2 S18198	notch protein homo
11	73.5	12.7	660	2 F70101	primosomal protein
12	73	12.6	1049	1 S19421	Anp-dependent perm
13	73	12.6	2437	2 S42612	transmembrane prot
14	73	12.6	861	2 A48825	Notch homolog Motc
15	73	12.6	1700	2 S08167	Balbani ring 3 pr
16	72	12.4	2318	2 S45306	notch 3 protein -
17	72	12.4	1111	2 B44018	laminin B2t chain
18	72	12.4	1193	2 A44018	laminin B2t chain
19	71.5	12.3	519	1 YRHU2	dopachrome Delta-i
20	71.5	12.3	2703	2 A24420	notch protein - fr
21	71.5	12.3	105	2 S23061	chorion protein -
22	71	12.2	1134	1 JN0711	protein-tyrosine k
23	71	12.2	396	1 TRYX34	alpha-lytic protei
24	71	12.2	341	2 D48435	cysteine proteinas
25	71	12.2	72	2 S39418	metallothionein 10
26	70	12.0	2907	2 A52728	fibrillin-2 precur
27	70	12.0	2871	2 A55624	fibrillin-1 precur
28	69.5	12.0	346	2 S34165	keratin, type II -
29	69.5	12.0	384	2 S64735	retrovirus-related
30	69.5	12.0	2524	2 A35844	Xotch protein - Af
31	69	11.9	2718	2 A23475	G surface protein
32	69	11.9	447	2 A39321	mucin - rat (fragm
33	68.5	11.8	540	2 B47417	insulin receptor-r
34	68	11.7	1138	1 S24066	protein-tyrosine k
35	68	11.7	388	2 S34672	alkaline serine pr
36	68	11.7	72	2 S39416	metallothionein 10
37	67	11.5	1136	1 S57845	protein-tyrosine k
38	67	11.5	1801	1 MWRTS	laminin beta-2 cha
39	67	11.5	72	2 S39417	metallothionein 10

ALIGNMENTS

RESULT 1  
A34372  
complement C6 precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 08-Jun-1990 #sequence\_revision 08-Jun-1990 #text\_change 07-Aug-1998  
C:Accession: A34372; A34235; A32109; A31894; A53072  
R:Haefliger, J.A.; Tschoop, J.; Vial, N.; Jenne, D.E.  
J. Biol. Chem. 264, 18041-18051, 1989  
A:Title: Complete primary structure and functional characterization of the sixth comp  
A:Reference number: A34372; MUID:90036879  
A:Accession: A34372  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-934 <HA>  
A:Cross-references: GB:J05064; NID:gl79703; PID:gl79704  
R:DiScipio, R.G.; Hugli, T.E.  
J. Biol. Chem. 264, 16197-16206, 1989  
A:Title: The molecular architecture of human complement component C6.  
A:Reference number: A34235; MUID:89380223  
A:Accession: A34235  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-118, 'E', 120-934 <DIS>  
A:Cross-references: GB:J05024; NID:gl78724; PID:9307228  
R:Chakravarti, D.N.; Chakravarti, B.; Parra, C.A.; Muller-Eberhard, H.J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 2799-2803, 1989  
A:Title: Structural homology of complement protein C6 with other channel-forming prot  
A:Reference number: A32109; MUID:89202413  
A:Accession: A32109  
A:Molecule type: mRNA  
A:Residues: 1-118, 'E', 120-491 <CH2>  
A:Cross-references: GB:J04506; NID:gl618465; PID:gl618466  
R:Chakravarti, D.N.; Muller-Eberhard, H.J.  
J. Biol. Chem. 263, 18306-18312, 1988  
A:Title: Biochemical characterization of the human complement protein C6. Association  
A:Reference number: A31894; MUID:89054009  
A:Accession: A31894  
A:Molecule type: protein  
A:Residues: 22-41 <CHA>  
R:Hobart, M.J.; Fernie, B.; DiScipio, R.G.  
Biochemistry 32, 6198-6205, 1993  
A:Title: Structure of the human C6 gene.  
A:Reference number: A53072; MUID:93291175  
A:Accession: A53072  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: G56G5G, 1-47, 'A', 49-118, 'E', 120-561, 'A', 563-618, 'A', 620-700, 'A', 702-763  
A:Note: sequence extracted from NCB1 backbone (NCBIP:134071)  
C:Genetics:  
A:Gene: GDB:C6  
A:Cross-references: GDB:119045; OMIM:217050  
A:Map position: 5p13-5p13  
C:Superfamily: complement c6; agrin inhibitor-like repeat homology; complement factor  
gy  
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-934/Product: complement C6 #status experimental <MAT>  
F:22-79/Domain: thrombospondin type 1 repeat homology <THRI>  
F:80-134/Domain: thrombospondin type 1 repeat homology <THR2>  
F:140-173/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:521-552/Domain: EGF homology <EGF>  
F:564-612/Domain: thrombospondin type 1 repeat homology <THR3>

F:644-699/Domain: complement factor H repeat homology <FH01>  
F:704-761/Domain: complement factor H repeat homology <FH02>

Query Match 14.8%; Score 86; DB 2; Length 934;  
Best Local Similarity 29.4%; Pred. No. 0.92;  
Matches 30; Conservative 10; Mismatches 36; Indels 26; Gaps 7;

QY 22 CSMEERAINERIQEVAGSLIFRAISSIGLCQSVTSRGDIATCPRGFAVTG---CTCGSAC 78

Db 699 CQTECIKPVQEVLTITPFRLYRIG---ESIE-----LTCPKGFVAGPSRYTCQG-- 749

QY 79 GSW--DVRAETTC-----HCOCAGMDWTGRCRCVOP 108

Db 749 NSWTPPISNLTCEKDTLTKLKGHCQ-LGQKQSGSEICMSP 789

## RESULT 2

S78549  
notch3 protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 17-Mar-1999

C:Accession: S78549; S71825

R:Joutel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <J01>

A:Cross-references: EMBL:U97669; NID:g2668591; PTD:g2668592

R:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabrier, H.; Mouton, P.; Alamowicz

x, M.M.; Weissbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.

Nature 383, 707-710, 1996

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MUID:97032728

A:Accession: S71825

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <J02

A:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C:Keywords: tandem repeat; transmembrane protein

F:318-349/Domain: EGF homology <EGF>

F:1838-1870/Domain: ankyrin repeat homology <AN1>

F:1871-1903/Domain: ankyrin repeat homology <AN2>

F:1905-1937/Domain: ankyrin repeat homology <AN3>

F:1938-1970/Domain: ankyrin repeat homology <AN4>

F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 14.1%; Score 82; DB 2; Length 2321;  
Best Local Similarity 29.9%; Pred. No. 4.8;  
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLECSVTSRGDIATCPRGFAVTGCTCGSACGSDVRAETTC----- 90

Db 1300 VGPVCCQ-TPRGPRACPPG--LSGFSRSPGCPGASNAACAAAPCLHGSGCRPAPLA 1356

QY 90 ---HCOCAGMDWTGRC 103

Db 1357 PFFRCACA-QGWTGPRC 1372

## RESULT 3

KXBO

Protein C (activated) (EC 3.4.21.69) precursor - bovine (fragment)

N:Alternate names: autoprothrombin IIA; plasma protein C

C:Species: Bos primigenius taurus (cattle)

QY 18 SKYLCSEERAINERIQEVAGSLIFRAISSIGLCQSVTSRG--DLATCPRGFAVTG---- 72

Db 57 SEEVCFEEA-REIFQNTEDTMAFWSFYSDQDQCRPSGSPCDLPCCGKCIDGLGGF 115

Query Match 13.9%; Score 80.5; DB 1; Length 456;  
Best Local Similarity 24.8%; Pred. No. 1.7;  
Matches 26; Conservative 10; Mismatches 40; Indels 29; Gaps 4;

Metch A protein - mouse (fragment)  
 N:Alternate names: Notch homolog  
 C:Species: Mus musculus (house mouse)  
 C:Date: 21-Jan-1994 #sequence\_revision 05-Jan-1996 #text\_change 14-Aug-1998  
 C:Accession: B49175; PH1569; S32109  
 R:Lardelli, M.; Lendahl, U.  
 Exp. Cell Res. 204, 364-372, 1993  
 A:Title: Metch A and Metch B--two mouse Notch homologues coexpressed in a wide variety of tissues  
 A:Reference number: A49175; MUID:93178563  
 A:Accession: B49175  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-387 <LAR>  
 A:Cross-references: EMBL:X68278; NID:g287987; PID:g287988  
 A:Experimental source: embryo  
 A:Note: sequence extracted from NCBI backbone (NCBIP:126159)  
 C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.  
 C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homolog  
 F:27-58/Domain: EGF homology <EGF>

Query Match 13.0%; Score 75.5; DB 2; Length 387;  
 Best Local Similarity 29.1%; Pred. No. 4.2;  
 Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

QY 48 GLEQCSV-----TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84  
 Db 141 GRCESVINGCRGPKCKNGGVCNAVNTARGFCRCPAGE--EGATCENDARTCGSLRCL 198  
 QY 85 AETTC-----HCQAGMDWTGARC 103  
 Db 199 NGGTCISGRSPFTCLCLG-SFTGPEC 223

RESULT 7  
 A35672  
 crumbs protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 21-Sep-1990 #sequence\_revision 18-Nov-1992 #text\_change 14-Aug-1998  
 C:Accession: A35672  
 R:Repass, U.; Theres, C.; Knust, E.  
 Cell 61, 787-799, 1990  
 A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila  
 A:Reference number: A35672; MUID:90263104  
 A:Accession: A35672  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2139 <TEP>  
 A:Cross-references: GB:M33753  
 A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for res  
 C:Genetics:  
 A:Gene: FlyBase:crb  
 A:Cross-references: FlyBase:FBgn0000368  
 C:Superfamily: EGF homology  
 C:Keywords: transmembrane protein  
 F:691-722/Domain: EGF homology <EGF>

Query Match 13.0%; Score 75.5; DB 2; Length 2139;  
 Best Local Similarity 25.0%; Pred. No. 18;  
 Matches 24; Conservative 10; Mismatches 37; Indels 25; Gaps 4;

QY 16 VSKTLCSEEAINEIRIQEAVGSLIFRAISSIGLEQCSVTSRGDLATCPRGFAVTGCTCG 75  
 Db 346 VNIHPLCQNPCLNNGACVVIG-----GSGALTCE-----CPKGAVARCEVD 388  
 QY 76 S-ACGSWDVRAETTC-----HCQAGMDWTGARC 103  
 Db 389 TDECAQQPQNGNSCIDRINGFSCDCSGTGYTGAF 424

RESULT 8

A;Title: A homolog of Drosophila Notch expressed during mammalian development  
A;Reference number: S18188; MUID:92111383  
A;Accession: S18188  
A;Molecule type: mRNA  
A;Residues: 1-2531 <WEI>  
C;Cross-references: EMBL:X57405; NID:g57634; PID:g57635  
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology  
F;1917-1949/Domain: ankyrin repeat homology <AN1>  
F;1950-1982/Domain: ankyrin repeat homology <AN2>  
F;1984-2016/Domain: ankyrin repeat homology <AN3>  
F;2017-2049/Domain: ankyrin repeat homology <AN4>  
F;2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match	12.7%	Score 74;	DB 2;	Length 2531;
Best Local Similarity	25.7%;	Pred. No. 29;		
Matches	36;	Conservative	12;	Mismatches 52; Indels 40; Gaps 8

  

Qy	4	LCALLLPVL---	GLLVASKTLCSMEEALNERIQEAVGSLIFRAISS--IGLEQCS-----	54
			:	:
Db	9	LCITLLPALAARGURCSQPS---	GTCLNGGRCEVANGTCAVCSGAFVGRQCQDPSPCL	64
Qy	54	-----VTSRGDL---	ATCPRGFAVTGC--TCGSAC-----GSWDVRAETT	88
			:	:
Db	65	STPCKNAGTCVVDHGGIVDACSPLGSGFLCLTPLANA	CLANPCRNCGTCDLLTLE	124
Qy	89	CHQOCAGMDWTGARCCRVQP	108	
Db	125	YKRCRP--PGWSGKSCQOADP	143	

Db 125 YKCRCP-PGWSGSKCQADP 143

RESULT 11

F70101

primosomal protein N (priA) homolog - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 18-Sep-1998

C;Accession: F70101

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, son, D.; Peterson, J.; Kertlage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Ha Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943

A;Accession: F70101

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-660 <KLE>

A;Cross-references: GB:AE001115; GB:AE000783; NID:g2687879; PID:g2687882; TIGR

A;Experimental source: strain B31

C;Superfamily: DEAD/H box helicase homology

F;158-519/Domain: DEAD/H box helicase homology <DEAD>

F;158-165/Region: nucleotide-binding motif A (P-loop)

[illegible]

Db 382 CSFGLIYHKENKLLCHYCSYKTKTASHCPQCESKD 417

RESULT 12

S19421

Arp-dependent permease ADP1 precursor - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YCR011c; protein YCR105  
C:Species: *Saccharomyces cerevisiae*

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 17-Jul-1998

C:Accession: S19421; S40914  
R:Goffeau, A.; Purnelle, B.; Skala, J.  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19420  
A:Accession: S19421  
A:Molecule type: DNA  
A:Residues: 1-1049 <GOF>  
A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264475; PID:g1907154; MIPS:YCR011c  
R:Purnelle, B.; Skala, J.; Goffeau, A.  
Yeast 7, 867-872, 1991  
A:Title: The product of the YCR105 gene located on the chromosome III from Saccharomyces  
A:Reference number: S40914; MUID:92160395  
A:Accession: S40914  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1049 <PUR>  
R:Skala, J.; Purnelle, B.; Goffeau, A.  
Yeast 8, 409-417, 1992  
A:Title: The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of  
K genes.

A:Reference number: S25353; MUID:92327849

A:Contents: annotation

C:Genetics:

A:Gene: SGD:ADP1

A:Cross-references: SGD:S0000604; MIPS:YCR011c

A:Map position: 3R

C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology

C:Keywords: ATP; glycoprotein; P-loop; transmembrane protein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-104/Product: ATP-dependent permease ADP1 #status predicted <MAT>

F:26-324/Domain: extracellular #status predicted <EXT>

F:325-341/Domain: transmembrane #status predicted <TM1>

F:406-607/Domain: ATP-binding cassette homology <ABC>

F:423-430/Region: nucleotide-binding motif A (P-loop)

F:550-557/Region: nucleotide-binding motif B

F:794-810/Domain: transmembrane #status predicted <TM2>

F:829-845/Domain: transmembrane #status predicted <TM3>

F:878-894/Domain: transmembrane #status predicted <TM4>

F:909-925/Domain: transmembrane #status predicted <TM5>

F:938-954/Domain: transmembrane #status predicted <TM6>

F:1025-1041/Domain: transmembrane #status predicted <TM7>

F:50.114,165,221/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:429/Binding site: ATP (Lys) #status predicted

Query Match 12.6%; Score 73; DB 1; Length 1049;

Best Local Similarity 35.9%; Pred. No. 17;

Matches 23; Conservative 4; Mismatches 21; Indels 16; Gaps 6;

QY 50 EQQVTSRGDLATCPRGFAVTGCT---CG-----SAGGSWD--VRAET-TCHQCAGMDWT 99

DB 71 ECNSTIGR---CECIEGFGAGDDCSPLCGGLSDSGNKRPIRAQNDTCHCD---NGWG 124

QY 100 GARC 103

DB 125 GINC 128

RESULT 13

S42612

transmembrane protein precursor - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 10-Jul-1998

C:Accession: S42612

R:Bierkamp, C.; Campos-Ortega, J.A.

Mech. Dev. 43, 87-100, 1993

A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern

A:Reference number: S42612; MUID:94128602

A:Accession: S42612

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2437 <BIE>

A:Cross-references: EMBL:X69088; NID:g433866; PID:g433867

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology

F:1915-1947/Domain: ankyrin repeat homology <AN1>

F:1948-1980/Domain: ankyrin repeat homology <AN2>

F:1982-2014/Domain: ankyrin repeat homology <AN3>

F:2015-2047/Domain: ankyrin repeat homology <AN4>

F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 12.6%; Score 73; DB 2; Length 2437;

Best Local Similarity 30.6%; Pred. No. 35;

Matches 22; Conservative 3; Mismatches 21; Indels 26; Gaps 5;

QY 51 CQSVTSRGD-----LATCPRGFAVTGCTC-----GSACGSWDVRAETTHC 91

DB 1028 CQNGSCQDGYGYKCTCPHY--TGLNCQSLVRWCSSPKKNGSC--WQOGASFTQC 1083

QY 92 QCAGMDWTGARC 103

DB 1084 ---ASGWTGIYC 1092

RESULT 14

A48825

Notch homolog Notch protein - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 14-Aug-1998

C:Accession: A48825

R:Reaume, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.

Dev. Biol. 154, 377-387, 1992

A:Title: Expression analysis of a Notch homologue in the mouse embryo.

A:Reference number: A48825; MUID:93050801

A:Accession: A48825

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-861 <REA>

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:119144)

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

F:26-57/Domain: EGF homology <EGF>

Query Match 12.6%; Score 73; DB 2; Length 861;

Best Local Similarity 30.3%; Pred. No. 14;

Matches 20; Conservative 5; Mismatches 21; Indels 20; Gaps 5;

QY 48 GLEQCQSVTSRGDLATCPRGFAVTGCTGCSACGSWDVRAETTHCQCAGMDWTGARC---- 104

DB 264 GLNCQNLVRWCDSAPCKN-----GGRC--W--QTNTQYHCECRS-GWTGVNCDVLS 309

QY 104 --CRVQ 107

DB 310 VSCEVK 315

RESULT 15

S08167

Balbani ring 3 protein - midge (Chironomus tentans)

C:Species: Chironomus tentans

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 20-Mar-1998

C:Accession: S08167

R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.

J. Mol. Biol. 211, 331-349, 1990

A:Title: The Balbani ring 3 gene in Chironomus tentans has a diverged repetitive str

A:Reference number: S08167; MUID:90172404

A:Accession: S08167

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1700 <PAU>

A:Cross-references: GB:X52263; NID:g7057; PID:g7058

C:Genetics:

A:Gene: BR3

A:Map position: 4



Query Match 12.6%; Score 73; DB 2; Length 1700;  
 Best Local Similarity 32.6%; Pred. No. 26;  
 Matches 14; Conservative 7; Mismatches 12; Indels 10; Gaps 2;  
 QY 62 TCPRGFV---TCCTC-----GSACGSWDVRAETTCQCA 94  
 || :||| :||| || : : ||| :|||  
 Db 1508 TCKQGFSPKSGCKILECNKKDPGCGAKKIWCQETCKCECA 1550

Search completed: October 2, 1999, 10:41:34  
 Job time: 176 sec